

Painful transition at the NIH

Elias Zerhouni has a mixed track record as director of the world's largest research agency — but the thrust of his reform effort should be supported.

Leading the National Institutes of Health (NIH) under the presidency of George W. Bush was always going to be a challenging assignment. That much was clear four years ago this week, when Elias Zerhouni, an Algerian-born radiologist with a strong administrative track record at Johns Hopkins University, was confirmed in the position by the US Senate.

Zerhouni has, on the whole, handled this potentially poisoned chalice adroitly (see page 17). But this year, as three years of flat budgets begin to bite, Zerhouni's tenure at the NIH is being openly attacked by some scientists. The focus of their ire is his 'Roadmap', a set of activities that run across different NIH institutes and attempt to implement Zerhouni's vision for the agency. The critics say that the Roadmap isn't working and is diverting resources and attention from basic scientific research.

There are three obvious responses to this criticism. The first is that the Roadmap expresses priorities for the NIH — including the more effective translation of basic research results into clinical practice — that most biomedical researchers would support and that any NIH director would need to implement at this time. The second is that it is not Zerhouni's fault that he is presiding over a period of flat budgets at the NIH, after the historic doubling of the agency's budget from 1998 to 2003. The third is that, in light of that doubling, public attacks on the Roadmap by what is now a rather well-funded community of biologists are going to make that community look bad.

Since its publication in 2003, the Roadmap has been the defining tenet of Zerhouni's directorship. By pushing for more interdisciplinary research, closer cooperation between the NIH's 27 historically independent research institutes and the rapid application of research results in the clinic, it addresses clear deficiencies in the NIH's previous performance. It also takes the agency in a direction that was called for by the Institute of Medicine in a landmark 2003 report, 'Enhancing the vitality of the National Institutes of Health'.

Funding for Roadmap activities is relatively modest — about \$400 million annually in a \$28-billion agency. But its influence is more pervasive than that number would suggest, and that is as it should be. Basic molecular biology will remain a critical component of NIH work but it will take place within a broader mandate of applying research to the good of public health. That is essentially what Zerhouni is trying to achieve, and his efforts deserve support.

Bush's man

Some of Zerhouni's problems with NIH scientists originate not in his Roadmap goals, but rather with his being an appointee of President Bush. He arrived when the NIH, despite its budgetary expansion, was chafing under a number of centralization efforts led by the then health secretary, Tommy Thomson, which were rightly perceived as threatening the agency's autonomy and credibility.

Although Thomson has gone, the NIH director has not been able

to wrest back the independence (to put out their own press releases, for example, or send who they like to scientific meetings) to which the NIH institutes are accustomed. NIH scientists might support Zerhouni more enthusiastically if they believed he had defended the agency's traditional autonomy as vigorously as possible.

Another important challenge for Zerhouni has been to explain in straightforward terms to Congress what the NIH has done with its increased budget. With the exception of the Roadmap, there's not much evidence that the strategic opportunity presented by the budget's doubling has been fully grasped.

The doubling took place during a leadership transition from Harold Varmus, through the prolonged interim directorship of Ruth Kirschstein, to the period when Zerhouni was finding his feet. As a result, there was no clear and consistent strategy for using the money. Individual institutes got their piece and moved on. A particular concern was that not enough young scientists were winning first-time grants during the most creative period of their careers. Zerhouni can't tell the institutes who to give grants to but he could, for example, have published data that would have exposed the failure of some institutes in this regard.

"NIH scientists might support Zerhouni more enthusiastically if they believed he had defended the agency's autonomy as vigorously as possible."

Surviving scandal

The other big event of Zerhouni's tenure has been a conflict-of-interest scandal that racked the NIH's main campus at Bethesda, Maryland, during 2004 and led to the introduction of stringent new rules governing its scientists' interaction with business. These rules are unpopular, but Zerhouni had to come down hard after congressional investigators unearthed a trove of commercial interactions of which the NIH itself had no record. That being said, a leader who inspired more loyalty and devotion from his troops might have won more understanding of what had to be done.

Partly as a result, the Bethesda campus — now probably the largest single research facility in the world — is not a particularly happy place. Post-2001 security measures have further battered its collegial ambiance, and back-biting is seldom far beneath the surface.

Improving coordination between the institutes remains a difficult issue for any NIH director. On his departure, Harold Varmus floated ideas for radical reorganization (see *Nature* **418**, 572; 2002) but, given the powerful backers for each institute's autonomy and the sheer inertia of US federal government, nothing has come of these.

The Roadmap makes the director's office more powerful than before, and that was a necessary change. On balance Zerhouni is doing a demanding job reasonably well. But he needs to work harder at persuading his staff and grantees that he is on their side. ■



Small first step

France's technology plans hold modest promise.

The European Union has identified targets for industrial research and development (R&D) as critical to its future, but is manifestly failing to meet them. Member states agreed at a 2000 summit in Lisbon to boost total R&D spending to 3% of gross domestic product by 2010 — an unrealistic target that the continent will not even begin to move towards unless it sharply boosts private-sector investment.

Last week, French president Jacques Chirac took a step towards addressing that challenge. He announced a €600-million (US\$760-million) plan, to be funded roughly 50–50 by the French government and by European companies, to support six technology projects. France would eventually increase this to more than €1 billion annually for dozens of similar projects, he said, in a bid to boost European innovation.

The projects will draw together research programmes on various topics, each championed by a single company, but with the work shared out between a network of others, and also involving laboratories at public research agencies and universities. The largest sum, €250 million, will fund a Franco-German programme called Quaero, led by Thomson, to develop multimedia search engines. The other projects are smaller, each receiving less than €100 million.

It is naive to suppose that whole new industries can arise from entrepreneurship alone. Governments can, and must, help create the conditions for innovation, from ensuring a skilled workforce and adequate research infrastructure, to tax breaks and other incentives. Even the United States has long nourished technology development through military contracts and national strategic programmes.

This logic is all the more compelling in Europe, where the reality is that students' and scientists' first thoughts before breakfast are not to go into a garage and launch a company on Nasdaq. The European

Commission's competitiveness ministry will be watching to ensure that the projects do not flout rules on subsidies. Its research ministry, although unhappy at being sidestepped, nonetheless welcomes what it sees as a long-overdue decisive response by member states to the critical lack of private-sector funding of research.

Nevertheless, there are reasons to be cautious about this initiative's prospects for success. France's industrial policies have a decent track record of picking winners: its state-sponsored 'grandes programmes technologiques' made the country a world pioneer in high-speed trains and nuclear reactors, and broke US domination of the civilian aerospace industry, creating Airbus and Ariane. These sectors also played to the strengths of France's centralized technocracy in organizing large national and strategic sectors.

But such strengths are less well suited to the dynamics of some of today's key industries — particularly information technology. Although Quaero has wisely ruled out trying to compete head-on with Google, focusing instead on research on next-generation web searching, a top-down approach is unlikely to match the flexibility and speed of bottom-up entrepreneurship.

French domestic politics will also affect the outcome of the new scheme. The 'grandes programmes technologiques' enjoyed national support by successive administrations whatever their political bent. In contrast, the new scheme is very much a *lapin* pulled out of a hat by Chirac in the decrepit dusk of two terms as French president, where research and innovation have otherwise been largely neglected.

Chirac's initiative at least has the merit of acknowledging the fact that European governments need to address the thorny question of their role in tackling the continent's industrial R&D deficit. ■

"The new scheme is very much a *lapin* pulled out of a hat by Chirac in the decrepit dusk of two terms as French president."

Amateur night

Ending an inhumane punishment.

Can the death penalty by lethal injection, as practised in 37 US states, be outlawed as inhumane? The courts are considering this question (see page 8), which can be separated from the more familiar one about the morality of capital punishment.

Their judgements are complicated by the fact that, since the procedure came into use three decades ago, few physicians, nurses or scientists have had anything to do with it. Physicians and nurses are ethically barred from assisting. Yet it is the dominant method in the United States, and has been used to kill more than 800 prisoners there.

Some information has emerged on the workings of the three-drug injection, however. One analysis (L. G. Koniaris *et al. Lancet* **365**, 1412–1414; 2005) suggests that 43% of prisoners might still be conscious, although totally paralysed, when the potassium chloride rips through their veins on the way to stopping their heart.

There are powerful arguments for abandoning the death penalty, regardless of its morality. DNA, for example, has helped prove many

death-row inmates innocent, exposing the flaws of an irreversible sanction. And statistical analyses indicate that the death penalty is disproportionately administered to minority populations.

Earlier this year, a California court told state authorities that they must persuade an anaesthetist to oversee an execution, come up with a new protocol for lethal injections — or face a hearing on whether the punishment is inhumane. The last option now looks likely.

If suitably qualified individuals refuse to help prepare a new protocol, the state will face the prospect of continuing to use amateurs to kill people with arbitrary and outmoded technology.

Scientists often abjure political activity, and could in this case argue that they are merely providing a basis from which policy-makers can make decisions. But this decision must be taken by the physicians and scientists themselves. All that is required is a refusal to participate. Men and women of science and medicine should stand shoulder to shoulder on this. Don't advise, don't prescribe, don't inject. Let the death penalty die a natural death. ■

"DNA has helped prove many death-row inmates innocent, exposing the flaws of an irreversible sanction."

RESEARCH HIGHLIGHTS

A tale of betrayal

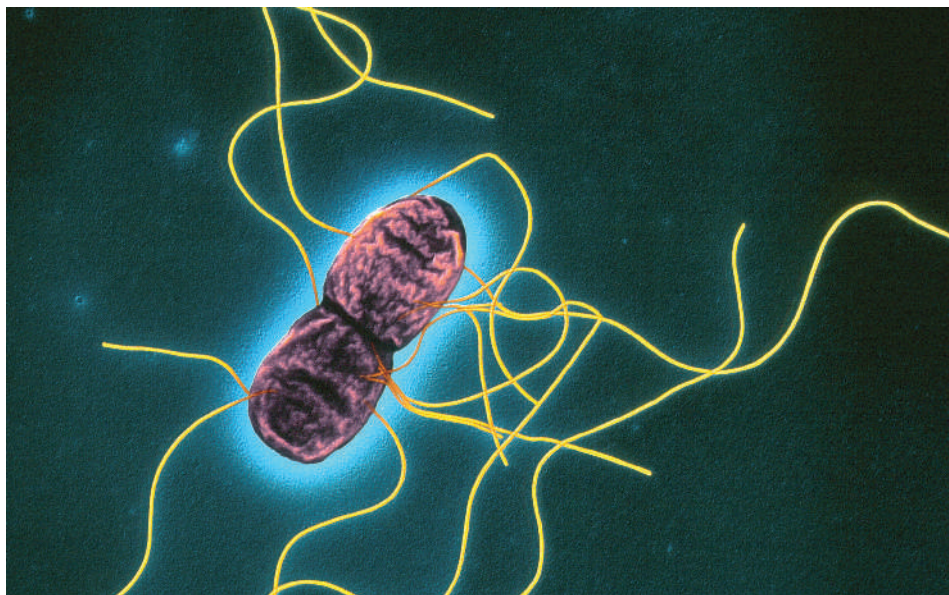
Nature Immunol. doi:10.1038/ni1336;

10.1038/ni1344; 10.1038/ni1346 (2006)

The flagellin protein that bacteria use to make their propeller-like filaments can betray the bugs when they infect cells. A trio of papers unpicks how this happens to *Salmonella typhimurium* (pictured).

It was already known that the host cell receptor 'Toll-like receptor 5' detects flagellin to trigger an immune response. Now two groups in the United States have shown that a second part of the immune response to *Salmonella*, mediated by the receptor Ipaf, also depends on flagellin.

Researchers at the National Institute of Immunology in New Delhi, India, show that *Salmonella*'s production of flagellin is boosted by lipids in the host cell — which should make the bugs easier to detect.



EYE OF SCIENCE/SPL

MALARIA**Resistant mozzies**

Science **312**, 577–579 (2006)

Researchers have honed in on the genes that make some mosquitoes resistant to the malaria parasite *Plasmodium falciparum*.

A group led by the University of Minnesota's Kenneth Vernick fed malaria-infected blood to more than 1,000 *Anopheles gambiae* mosquitoes captured in Mali. After seven days, the researchers grouped the insects by the number of parasites in their mid-gut. By comparing the DNA of mosquitoes in the different groups, they traced resistance to a single region of the genome. They also used RNA interference to pinpoint one immunity gene, known as *APLI*, that controls parasite number.

The researchers speculate that mosquitoes are naturally resistant to malaria, and that a small number — which could be targeted as part of a control strategy — become susceptible through defects in their immune system.

NUCLEAR PHYSICS**Extending the elements**

Phys. Rev. Lett. **96**, 142502 (2006)

Where does the Universe get its heavy metals from? The formation of elements heavier than iron happens in supernovae, where atoms absorb extra neutrons and protons. But this doesn't explain the observed abundances of all elements, such as some proton-rich isotopes of ruthenium and molybdenum.

Carla Fröhlich of the University of Basel,

Switzerland, and her co-workers propose a new mechanism for how some of these elements form. They say that a supernova explosion produces a short-lived shell of proton-rich material. The protons aren't readily captured by large, proton-rich nuclei because of their mutual repulsion. But by combining with antineutrinos, which are also common in the debris, the protons may become neutrons. The neutrons are then easily absorbed by nuclei, building heavier elements.

POLAR SCIENCE**Wind of change**

Geophys. Res. Lett. **33**, L08605 (2006)

Fragmentation of sea ice may accelerate melting in the Arctic by strengthening wind-driven currents, say researchers in Japan, Canada and the United States.

During the late 1990s, sea-ice cover in the Pacific quadrant of the Arctic more than

halved. The sudden and dramatic melting, following a period of slow decline, was linked to an increased inflow of warm, summer water from the Pacific. But it was unclear what had changed, and why the ice did not recover.

Koji Shimada of the Institute of Observational Research for Global Change in Yokosuka, Japan, and his colleagues blame a novel feedback mechanism. Field research (pictured) shows that the gradually shrinking ice sheet had become more mobile. They suggest that this allowed the prevailing winds to set up currents that washed in the warm water, pushing the system over a threshold.

SYNTHETIC CHEMISTRY**Two more ways to Tamiflu**

J. Am. Chem. Soc. doi:10.1021/ja0616433;

10.1021/ja061696k (2006)

Two new methods have been devised to synthesize the anti-influenza drug oseltamivir.

Demand for the drug, also known as Tamiflu, has soared amidst fears that bird flu might mutate to trigger a human flu pandemic. The new syntheses avoid one of the problems that has limited production so far: a difficult-to-obtain starting material known as shikimic acid.

One reaction scheme, from E. J. Corey's group at Harvard University in Cambridge, Massachusetts, begins with the simple molecules butadiene and acrylic acid. The second, more complex synthesis, designed by Masakatsu Shibasaki of the University of Tokyo, Japan, and his colleagues, uses compounds known as *meso*-aziridines as a starting material.



C. LINDER/WOODS-HOLE OCEANOGR. INST.

IMMUNOLOGY

Baby's first bacteria

J. Exp. Med. **203**, 973–984 (2006)

When babies encounter bacteria in the birth canal, it may teach them that the body's bugs are harmless.

Mathias Hornef of the University Clinic of Freiburg and his colleagues show in mice that epithelial cells lining the gut react to a bacterial endotoxin and release inflammatory molecules before birth — but the same cells do not do so in newborns or adults.

The authors present evidence that endotoxin from the mother's microflora, possibly from the vaginal tract, passes through the mouths of babies. This briefly stimulates the gut cells and teaches them to tolerate certain bugs, something not seen in babies born by caesarean section. The process may help useful bacteria to set up shop in babies' intestines after birth.

MATERIALS SCIENCE

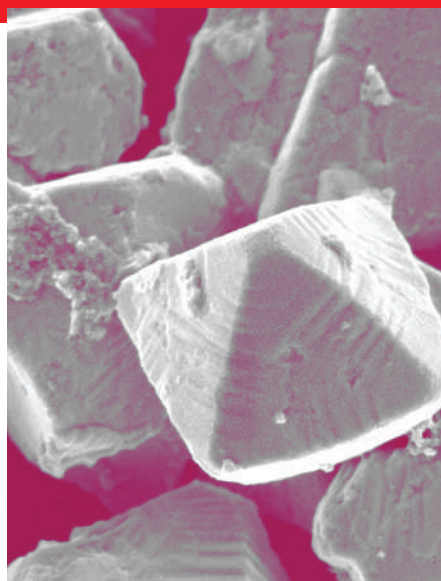
Modern alchemy

Science **312**, 420–424 (2006)

Materials scientists have come up with a modern-day twist to the alchemists' dream of turning base metals into gold. They have turned gold and silver into diamond — or at least, diamond-like crystals.

Bartosz Grzybowski of Northwestern University in Evanston, Illinois, and his colleagues started with an aqueous solution of gold or silver nanoparticles capped with a single layer of organic molecules. By evaporating the water from the solution, they grew crystals in which the particles were arranged like the carbon atoms in diamond. Binary superlattices containing both gold and silver nanoparticles also showed diamond-like packing.

The structure is unusual because the forces



between such particles generally make them close-packed: here, weak attractive interactions allow for a more spacious arrangement. Each micrometre-sized crystal (pictured) contains millions of nanoparticles.

MICROBIOLOGY

Parasitic rodeo

Cell **125**, 261–274 (2006)

Researchers have found that the common food-borne parasite *Toxoplasma gondii* uses a cowboy trick to rustle nutrients from its host.

Isabelle Coppens of Johns Hopkins University in Baltimore, Maryland, and her colleagues show that the parasite co-opts the host cell's microtubule network to lasso nutrient-rich lysosomes. The parasite wraps the lysosome in a protein called GRA7 before engulfing it. The role of the protein is not clear, but parasites engineered to lack it could not corral the lysosomes, leading to their death.

CHEMISTRY

No flop means less flip

J. Am. Chem. Soc. **128**, 5332–5333 (2006)

Thanks to the violation of mirror symmetry, or parity, in processes involving the weak nuclear force, the two enantiomers of a chiral

molecule — mirror-image molecules that cannot be superimposed — should have different vibrational energies. The difference is so tiny that it has never been detected, but Andrey Fokin of the Justus Liebig University in Giessen, Germany, and his co-workers are not ready to give up.

The researchers have made chiral molecules that they say are well suited to searching for the effect. The molecules have four different halogens at the corners of a carbon-cube (cubane) framework. The molecules' rigid framework would help to reduce noise in the measurement because, unlike the floppier molecules tested so far, the configuration is less likely to switch between enantiomers.

CELL BIOLOGY

The missing link

Cell **125**, 327–341 (2006)

A missing link in a key biochemical signalling pathway has been identified. The Wnt pathway is one of a handful of cell-to-cell signalling systems that control embryonic development and maintain adult tissues. Defective Wnt signalling is linked to cancer.

Konrad Basler and his colleagues at the University of Zurich in Switzerland studied how β -catenin, a component of the Wnt pathway, transfers signals from a cell's membrane to its nucleus. They identified an unexpected link in the chain. They show that the protein parafibromin — previously identified as a tumour suppressor in parathyroid glands — binds β -catenin as well as an RNA polymerase enzyme that transcribes genes.

Correction

A picture of a mouse blastocyst was incorrectly described as a one-cell embryo in 'Assume the position' (*Nature* **440**, 973; 2006).

B. A. GRZYBOWSKI

JOURNAL CLUB

Yoshinori Watanabe
University of Tokyo, Japan

A geneticist reveals an unexpected pattern of chromosome segregation.

Recent experiments have reshaped my ideas about how chromosomes behave in the two types of cell division — mitosis and meiosis.

Chromosomes in cells exist in pairs. Human cells have 23 pairs, for example. One chromosome in each pair is derived from dad, the

other from mum. When the cell divides, these chromosomes are copied and then shared out between the daughter cells.

Mitosis produces identical daughter cells — with one copy of each 'mum' chromosome and one of each 'dad'. I had thought, wrongly as it turns out, that the way mum and dad chromosomes divide in mitosis was always independent of each other.

Sometimes, genetic material is shuffled between the mum and dad chromosomes in a pair. This process is known as

recombination. Pentao Liu and his colleagues from the National Cancer Institute in Frederick, Maryland, showed that recombined chromosomes in mouse embryonic stem cells always segregated into different daughter cells (P. Lui *et al. Nature Genet.* **30**, 66–72; 2002), whereas I had supposed it should be random.

A further blow to my preconceived notions came from Athanasios Armakolas and Amar Klar of the National Cancer Institute (A. Armakolas & A. Klar *Science* **311**, 1146–1149; 2006). They found that

the segregation pattern of recombined chromosomes changes as the stem cells differentiate into more specialized cells.

This hints to me how meiosis, which I study, evolved from mitosis. In meiosis, the chromosomes are divided so that one daughter cell gets both mum chromosomes, the other both dads. Recombination, which may take place even during the repair of an injured chromosome, seems to have an intrinsic ability to shift mitosis towards splitting mums and dads, the hallmark of meiosis.

NEWS

US posts sensitive climate report for public comment

Climate experts have expressed surprise and concern about a US government decision to release a politically sensitive report when it is still in draft form.

Access to the first section of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report, due for publication in 2007, has been open for around a month. IPCC reports are designed to inform political negotiations on climate change, and key statements in the assessments can influence debate for years after publication. The first section, called Working Group I (WGI), offers an overview of the science of climate change and predictions about the possible course of global warming.

The US government says its move will allow the broadest range of experts to comment, and will avoid accusations that any groups have been denied access. But the decision effectively to publish the report worries some researchers, who say it could undermine what has previously been a confidential review process.

"If the US government is going to allow lobby groups, or persons associated with them, input into the government review, there is a serious chance that the whole thing could get hijacked," says one US government scientist involved in the process.

The WGI chapters of the assessment have completed one round of scientific review and are now with national governments. Most countries have solicited comments from a small number of experts, but the US Climate Change Science Program, which coordinates US climate research efforts on behalf of the government, posted the draft online on 7 April, together with a notification in the *Federal Register*. The climate science programme also sent out an e-mail to "thousands" of scientists, environmental groups and industry lobbyists, inviting them to comment. The website and report itself note that the contents should not be distributed or quoted, despite the fact that they can be accessed by asking for a password, which is provided automatically.

"I was quite surprised," says a senior IPCC scientist, who spoke on condition of anonymity. "I never thought they would do something like this." Rajendra Pachauri, the panel's chairman,

did not learn of the decision until after the report was posted online. He says that it is not in keeping with past practice, but adds that he could not interfere with how governments choose to collect comments. Lead authors on the WGI report declined to comment.

Roger Pielke Jr, a climate-policy expert at the University of Colorado, Boulder, suggests that the move could be seen as a deliberate strategy to defuse the newsworthiness of the final report. "If the report is already out there in circulation, then the 'news' value is likely to be much diminished when the official report is finally released," he says. The current US administration has been critical of the workings of the IPCC and its conclusions.

Up for debate

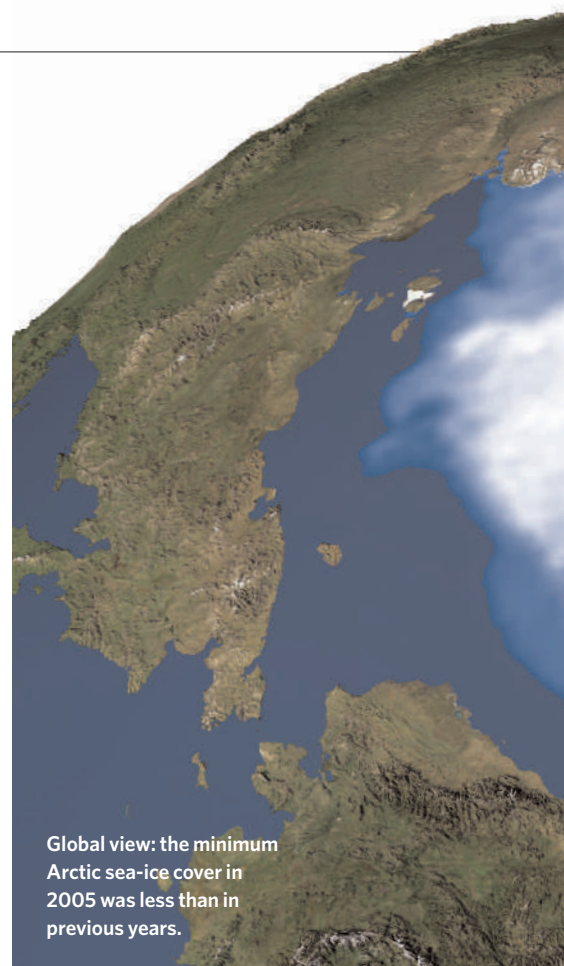
But Pielke adds that this isn't the only interpretation: "Less cynically, I do think that scientific assessments should be done much more in a continuous mode than the discrete approach favoured by the IPCC. So if this is a step in

that direction, there could be some value." Other climate experts contacted by *Nature* also welcomed the open approach, while questioning the motivation.

Governments have until 2 June to send comments to the IPCC. Those familiar with the process say that the chapters are unlikely to change substantially, but that the technical and policy-maker summaries will attract many comments. The current draft, which represents the message that the scientific authors want to present to policy-makers, contains few statements that will surprise climate researchers, but its tone is much more confident than that of its predecessor, published in 2001. And that, say researchers, will make it harder for sceptical politicians and lobbyists to attack climate predictions.

"People won't be punching holes in the science," says Jay Gulledge, a senior research fellow at the Pew Center on Global Climate Change in Arlington, Virginia. Emily Shuckburgh, a climate researcher at the University of Cambridge, UK, agrees: "If you're a sceptic, it's difficult to see where to attack on the modelling side."

One critical number in previous reports has been the sensitivity of the climate to increases in greenhouse-gas levels. In 2001, scientists esti-



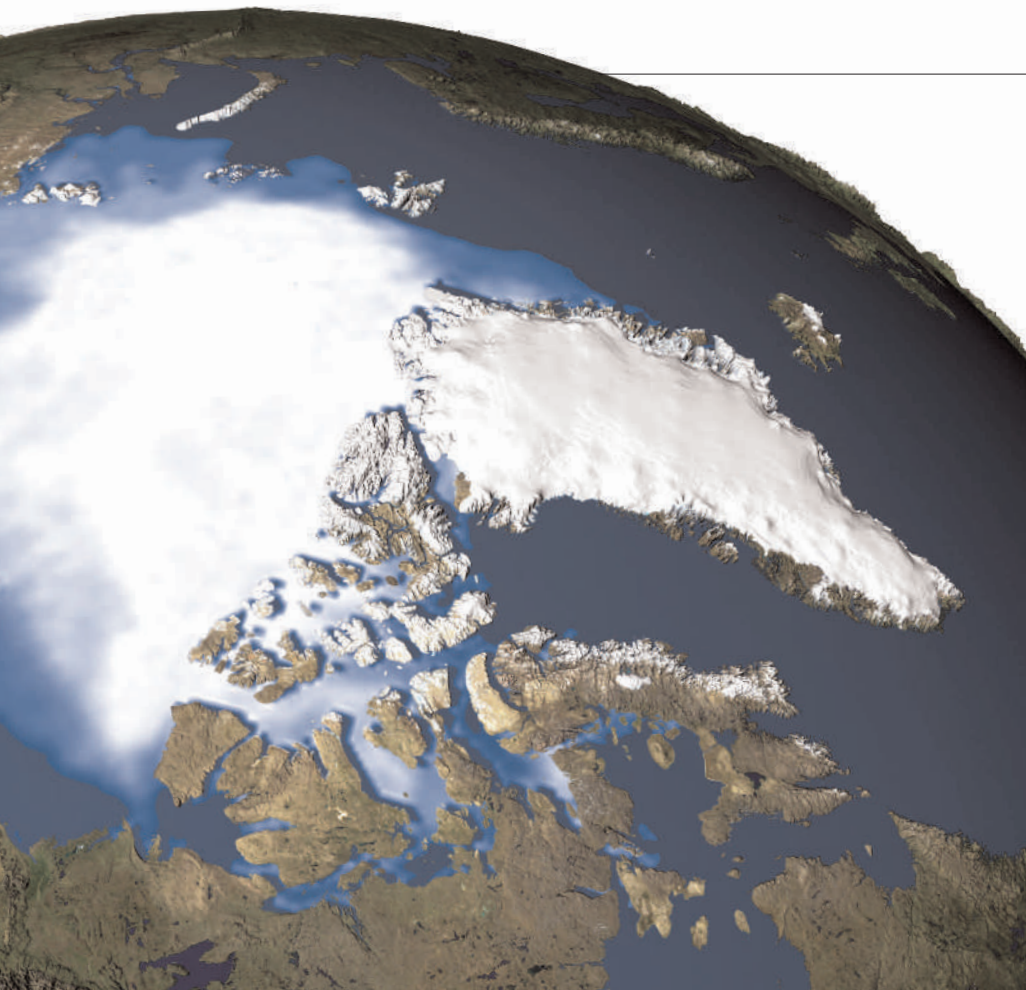
Global view: the minimum Arctic sea-ice cover in 2005 was less than in previous years.

mated that a doubling of carbon dioxide levels would cause an increase of 1.5–4.5 °C, but acknowledged that this range was little more than a best guess. The draft 2007 WGI report describes how new models and data sets allow the range to be properly quantified. It estimates the effect of doubling carbon dioxide as a rise of 2.0–4.5 °C and, for the first time, suggests a single most likely figure: 3 °C. This estimate is already widely accepted by climate scientists.

Commitment to change

Another set of predictions that have become much more robust are those about 'commitment' — the ongoing climatic changes that would be expected even if greenhouse-gas levels could be stabilized. The existence of commitment was acknowledged in the last WGI report, but no number was given in the policy-makers' summary. In contrast, the 2007 summary stresses that even if greenhouse gases level off now, warming will continue at about the current rate for several decades.

The error bars have also shrunk substantially on one of the biggest uncertainties in 2001 — the role of aerosols such as soot from fires, which exert a cooling effect by reflecting sunlight. In addition, certainty over politically



NASA

important statements about whether climate change has already been observed has increased. Data on twentieth-century changes in precipitation and sea-level rise are now more precise, and the risk of ocean acidification is detailed for the first time. Such assertions are likely to be seized on by environmental groups if they appear in the final document.

Reports of Working Groups II and III, which cover the consequences of climate change and the attempts to tackle it, are due later in 2007. An overall policy-makers' summary covering all three components will then be produced. In previous years this has been negotiated line by line by politicians, some of whom have been accused of seeking to downplay certain findings to protect national interests, such as the petroleum industry. The 2001 version took four days to finalize.

One graph in that summary, known as the 'hockey stick' and used to illustrate temperature change over the past millennium, has caused controversy for years, and is omitted in the current draft of the 2007 WGI policy-makers' summary, although an updated version is in the larger document. This is likely to be

seized on by sceptics as evidence that they were right to question the graph's validity; however, there is a similar graph in the current summary, which shows a dramatic twentieth-century rise in the degree to which greenhouse gases trap energy in the atmosphere.

If the confident tone of the draft survives, policy experts say it could have quite an impact. Kyoto Protocol negotiations have started to examine what will happen when the current agreement expires in 2012. Similar discussions are being run by the Group of Eight industrialized nations, and by the Asia-Pacific Partnership on Clean Development and Climate, which includes the United States and

Australia, neither of which is party to Kyoto.

"Several negotiation processes are ongoing," says Michael Grubb, a climate-policy expert at Imperial College, London, who is involved in WGII and WGIII. "That creates ripe ground for substantial political impact."

Jim Giles

Comments on the IPCC report can be submitted until 9 May. Details of how to access it can be found at

♦ www.climate-science.gov/Library/ipcc/wg14ar-review.htm

"If you're a sceptic, it's difficult to see where to attack on the modelling side."

ON THE RECORD

"Hedgehogs have been around for 20 million years — we want them around in the next century."

Ecologist Paul Bright expresses concern for Britain's disappearing hedgehogs. They will now be tracked by the Internet-based 'HogWatch'.

"Among the astronaut's needs are guidelines on performing prayers in space."

A Malaysian government official discusses unique concerns faced by the country's first astronaut, a Muslim.

Source: Reuters

SCORECARD

Space-age scavengers
NASA forms a 'roadkill posse' to deter vultures around the shuttle launch pad by removing the carrion on which they feed. The birds came under fire after one struck Discovery as it took off last year.

Clam confusion
Pollutants are sowing a gender crisis among Britain's bivalves, causing as many as 60% of male clams to make eggs.

Pie gobblers
Drug company Bayer, maker of the Alka-Seltzer heartburn remedy, drops its sponsorship of a worldwide speed-eating contest.

NUMBER CRUNCH

Biochemist Akira Endo has an obsession: "I love mushrooms and moulds." In 1973, he isolated the chemical ML-236B from a fungus — it was to form the basis of the famed cholesterol-lowering drugs known as statins. The work wins him this year's Japan Prize for new therapies.

6,000 mushrooms and moulds were studied by Endo before he discovered ML-236B.

Nothing was earned by Endo for statins, because his patent expired before the drugs took off.

¥50 million or US\$440,000 was awarded to Endo as a Japan Prize winner.

SPECIAL REPORT

Will medics' qualms kill the death penalty?

US judges are starting to acknowledge that lethal injection could be an excruciating way to die. **Emma Marris** investigates whether the medical community's refusal to assist could help end the practice.

Do no harm. Nearly all US MDs and PhDs recite this three-word oath instinctively when asked why they object to helping with executions. "It violates our ethical oath and erodes public trust," says Priscilla Ray, head of ethics at the American Medical Association (AMA), which prohibits members from participating in executions. Even helping to design a more humane protocol would disregard the AMA code, Ray says. "Formulating a way to kill somebody would violate the spirit of the policy."

As a result of that stance, lethal injections — the dominant method of execution in the United States — are generally carried out by technicians without scientific or medical training, and the protocol does not seem to have been reviewed in 30 years. Typically, prisoners are injected with three drugs in sequence meant to knock out, paralyze and then kill (see graphic). But many experts are concerned that because of that lack of training, the first drug does not always knock the condemned out properly, leaving him or her paralyzed but excruciatingly conscious as the lungs stop moving, and burning potassium chloride races towards the heart.

None of this seems until recently to have hindered the popularity of the technique. It was devised in 1977 by Oklahoma medical examiner Jay Chapman, despite his admitted lack of experience. It was meant to be more humane than the electric chair, which was occasionally setting inmates on fire, and the method has since been copied by 36 other states, apparently without any modification. It has killed more than 800 prisoners so far in the United States, and other countries, including China, Guatemala and the Philippines, have adopted the method in recent years. US prisoners have brought cases against their states for years, arguing that lethal injection is a "cruel and unusual punishment", and therefore prohibited by the constitution. But they have not made significant progress in the courts.

Now, however, the tide may be turning. In recent months, several death-row inmates and

their lawyers have had some successes, helped by evidence that as many as 43% of prisoners may be conscious until the point of death (L. G. Koniaris, T. A. Zimmers, D. A. Lubarsky and J. P. Sheldon *Lancet* 365, 1412–1414; 2005).

Last week for example, the Supreme Court heard arguments on whether condemned murderer Clarence Edward Hill can use a particular legal avenue to object to the way the state of Florida plans to execute him. Of the judges involved in the case, Justice Anthony Kennedy seemed upset by Florida's insistence on the standard protocol, asking: "Doesn't the state have some minimal obligation to do research to find a method as humane as possible?" If the Supreme Court lets Hill try this method of appeal in the district court, they may see it on their docket in the near future.

In California, executions have been suspended until a September hearing rules on the constitutionality of the method used there. The hearing could have been avoided had the state been able to fulfill a judge's order that murderer and rapist Michael Morales be exe-



cuted by a single dose of anaesthetic administered by a licensed practitioner, or that a trained anaesthesiologist be present. The state couldn't find any medical personnel willing to participate.

And in North Carolina last month, a judge ruled that the execution of Willie Brown Jr, a convicted murderer, could proceed only if a medical expert ensured Brown was fully anaesthetized when the final two drugs in the sequence were injected; anonymous medical personnel were persuaded to take part.

Ethical dilemma

These latest cases place doctors and scientists, who might be able to come up with an alternative method, centre-stage in the debate. "There is a real problem here," says Richard Dieter, executive director of the Death Penalty Information Center, a clearinghouse, an agency based in Washington DC with no official position on the death penalty itself. "Lethal injection is a medical procedure. I think that it almost requires the presence of the medical community."

Nathan Barankin, spokesman for the California attorney-general's office believes that arguing this need for trained overseers is the operational strategy of campaigns against the death penalty. "The goal of death-penalty opponents is to get a court order that says that lethal injections can only be administered by licensed professionals, because the ethics of medical

LETHAL INJECTION — THE DRUGS USED

1. SODIUM THIOPENTAL

This anaesthetic turns off the mind — or so prison officials assume. But because of obesity, prior drug use or incompetent administration, some inmates may not be rendered unconscious.

2. PANCURONIUM BROMIDE

Paralyzes the body. If not properly anaesthetized, prisoners may experience suffocation as their lungs stop, and be unable to show their awareness.

3. POTASSIUM CHLORIDE

Stops the heart. Said to be a searing, painful death if the person is conscious.



43% of executed prisoners may be conscious for the whole sequence, according to a study that looked at post-mortem concentrations of sodium thiopental (L. G. Koniaris, T. A. Zimmers, D. A. Lubarsky and J. P. Sheldon *Lancet* 365, 1412–1414; 2005).



Evidence that lethal injections are inhumane are being used to delay executions.

death penalty, vehemently disagree with Lubarsky and colleagues' finding that almost half of the executed prisoners had been awake for their deaths. They question the assumption that the level of anaesthetic post-mortem would be the same as at the time of death. The dispute continues.

The research, even though it only used post-mortem records, presented its own ethical challenges. Lubarsky says he worried that if the data showed lethal injection was knocking people out without fail, he would have two opposing ethical obligations: "Would we be scientifically ethically obligated to publish the results, which might go against physicians' ethics for aiding the death penalty?" He decided to publish whatever the outcome.

Punishment by pain

Lubarsky argues that more than voluntary non-participation is called for. "I believe that if death-penalty opponents were serious, they would seek the names of the physicians who participate under the freedom of information act," he says, in the hope that those doctors would then lose their licenses.

Many physicians who help with executions do so despite being opposed to the death penalty, however. Carlo Musso, a Georgia physician, is one of the few willing to be named. He works in jails and feels both that the death penalty is wrong and that his patients deserve a good death. Musso was reportedly summoned to Washington by the AMA as a result of going public in Gawande's article, and could be kicked out of the association.

Sara Tofte, an author of a report on lethal injection released last week by New York-based Human Rights Watch. She is among those who hope that if physicians and other experts refuse to touch the death penalty, it will be impossible for states to continue the practice. But she says she is disheartened by the research she did for the report: "I think a number of individuals say 'Who cares if the inmate feels a little bit of pain? If they murdered someone, maybe they deserve to feel pain.'"

The argument that lethal injection is inhumane "is coming from a scientific and legal perspective — it's not a popular groundswell coming from the public", says Dieter. He does not believe the stalemate will damage the death penalty. "Even if 100% of doctors refuse to participate, it will not be enough," he says. "The state will do it without them." ■

professionals prohibit them from participating."

Opinion among medics about the stance they should take — violate 'do no harm', or allow prisoners to die in pain? — is divided, however, and often not along predictable lines. Atul Gawande, an assistant professor of surgery at Harvard Medical School in Boston, Massachusetts, interviewed several doctors who help with executions for a March article in the *New England Journal of Medicine* (A. Gawande, *N. Engl. J. Med.* 354, 1221–1229; 2006).

Gawande supports the death penalty, but argues that doctors ought not use their skills to do anything but heal. "If I have to choose between core principles of medical ethics and the death penalty, I will choose medical ethics," he says. "If the only way we can do an execution and have it stand up to constitutional scrutiny is with physician involvement, then it is more important to have physicians not be involved and to lose the death penalty."

David Lubarsky, chair of the anaesthesiology department at the University of Miami in Florida, opposes capital punishment. He is a co-author of the *Lancet* study published last year, which analysed the amount of anaesthetic in the blood of 49 executed prisoners. Some other anaesthesiologists, also opposed to the

"Doesn't the state have some minimal obligation to do research to find a method as humane as possible?"



FLOOD OF THE DANUBE
What caused the flooding, and should we be expecting more?
www.nature.com/news

L. PARRETT

Tempers flare at hurricane meeting

MONTEREY, CALIFORNIA

Hurricanes may be becoming more intense because of global warming, but so is research on the topic.

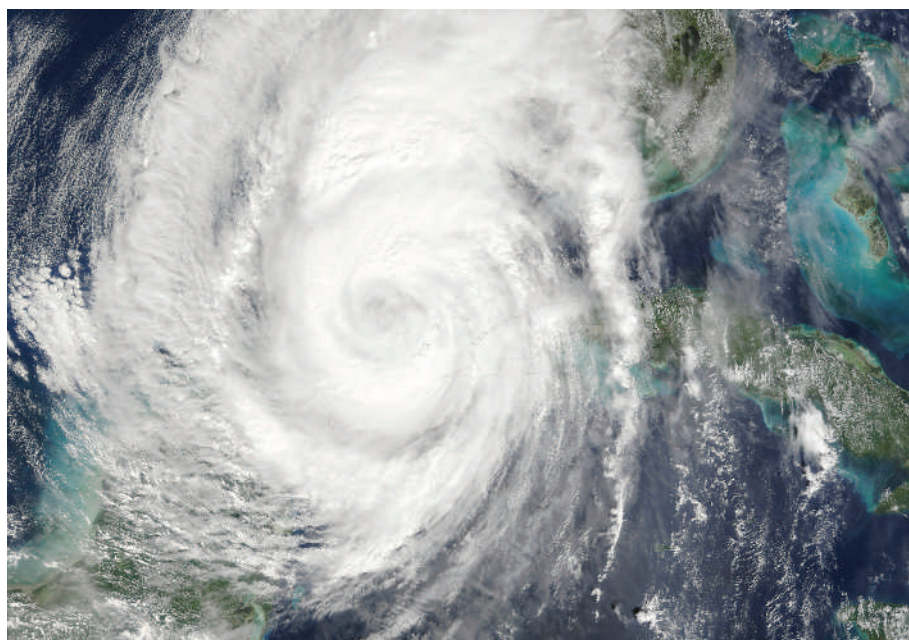
Last week, meteorologists gathered in Monterey, California — the meeting was to have been held in New Orleans — to hear new evidence supporting the proposed link between rising sea surface temperatures and more-powerful hurricanes. The fresh crop of research was triggered by two papers arguing for that connection, published by coincidence during the height of the devastating Atlantic hurricane season of 2005 (K. Emanuel *Nature* **436**, 686–688; 2005; P. Webster *et al. Science* **309**, 1844–1846; 2005).

The effects of global warming, many say, are already in evidence. “It’s not going to happen in the future — it is actually happening now,” says Greg Holland of the National Center for Atmospheric Research in Boulder, Colorado.

But not all are convinced, and there were raised voices in the hallway discussions at the American Meteorological Society meeting. At a lively panel discussion on 25 April, the moderator accepted audience questions only on written cards, and admonished the speakers not to indulge in personal attacks. This prompted Christopher Landsea, of the National Hurricane Center in Miami, to quip: “I get along personally with everyone in the field, and I want to continue that — even if they’re wrong.”

At issue is whether the historical record of cyclones is complete enough for accurate conclusions to be drawn about changes from past patterns. Many researchers called for the databases to be brought up to date by including modern assessments of past storms, including their intensities. It is a daunting task that, for now, is being done only for the Atlantic basin by Landsea and his colleagues.

Even given the gaps in the database, several new studies suggested that rising sea surface temperatures are having a noticeable effect on cyclones. Peter Webster of the Georgia Institute of Technology in Atlanta, co-author of one of last year’s papers, presented data hinting that not only are hurricanes growing more intense over time, but that the length of the storm season has increased as well. Starting from 1950, he told the meeting, the storm season has grown longer in the Atlantic by about five days per decade, in the northeastern Pacific by eight days per decade, and in the



Hurricane Wilma helped to make 2005 a record hurricane year, but is global warming to blame?

J. SCHWALTZ

northwestern Pacific by ten days per decade.

In Britain, researchers at the Benfield Hazard Research Centre in Surrey have run climate simulations suggesting that half the recent rise in hurricane activity in the North Atlantic can be explained by the observed increase in sea surface temperature in the region where these hurricanes develop. A warmer ocean would, in theory, provide more fuel for hurricanes to intensify.

And in Japan, a team has used the Earth Simulator supercomputer to run high-resolution simulations of global climate, both in today’s conditions and in a world warmed by higher levels of atmospheric carbon dioxide. Preliminary results suggest that, in the latter scenario, the number of tropical cyclones would drop by about 30% worldwide. But the number would rise in the Atlantic, and storm intensity would increase worldwide (K. Oouchi *et al. J. Meteorol. Soc. Jap.* **84**, 259–276; 2006).

Kerry Emanuel, a hurricane expert at the Massachusetts Institute of Technology, and author of the other paper published last year, says that many of these models need more work before firm conclusions can be drawn. But he spent his time as a panellist demolishing another popular tenet of climate research: that a natural temperature cycle known as the

Atlantic Multi-Decadal Oscillation (AMO) controls the formation of Atlantic storms during the peak of the hurricane season. Removing the oscillation would remove one of the last arguments that hurricane patterns are driven by a natural cycle.

“There may be such a thing as an AMO, but it’s not affecting the sea surface temperature in the North Atlantic in the late summer,” says Emanuel.

Few of these studies have appeared in peer-reviewed journals yet, and not everyone is persuaded by the conclusions. Johnny Chan, of the City University of Hong Kong, presented data suggesting that in the northwestern Pacific, numbers of typhoons do not track with rising sea surface temperatures.

One major critic is William Gray, a long-time forecaster at Colorado State University, Fort Collins, whose climate theories have been criticized by many researchers. He continued to insist that hurricanes are driven more by natural patterns of ocean circulation than by human-induced global warming. “I don’t care what anyone says — in the end the data will prove this to be true,” he says. “Why try to read all these demonic influences into the system?”

Forecasters, including Gray, are predicting that a busier-than-usual Atlantic hurricane season will start on 1 June.

Alexandra Witze

“Speakers were admonished and told not to indulge in personal attacks.”

Chemists get out begging bowl to avert closure

Chemists desperate to save their university department from closure have launched a £1.2-million (US\$2-million) fund-raising drive. The proposed closure is part of a growing trend in Britain that many say is triggered by insufficient funding for chemistry courses, and it has drawn condemnation from chemists around the world. But researchers at the University of Sussex in Brighton are fighting back.

They aim to persuade pharmaceutical companies to come up with the cash, which would fund three new posts in the department for five years. Gerry Lawless, head of chemistry at Sussex, hopes the pledges would be enough to convince officials at the university to keep the subject alive.

Although such enterprises are relatively common on US campuses, they are more unusual in Britain. "I think it's inherent in British culture that we don't like going cap in hand," says Lawless. "But that shouldn't be the case."

The crisis started when an investment strategy championed by the university's vice-chancellor, Alasdair Smith, became public in March this year. The controversial plan proposed to end teaching and research in physical and inorganic chemistry, and to create a new department of chemical biology.

"Sussex has not been accepting chemistry students at a high enough rate to maintain a broad chemistry department," explains Smith. The department's annual intake has stood at just 20–25 students for several years. Although applications are up by 40% this year, "converting applications into acceptances is by no means guaranteed," Smith says. The department has also lost five key staff to other universities in the past three years, reducing coverage of certain areas of chemistry and making it difficult to attract enough government funding for teaching. Smith says this will result in a substantial deficit if the situation continues.

But eminent chemists around the world have reacted with dismay, bombarding Smith with letters of stinging criticism (see 'Messages of rebuke'). Former Sussex chemist and Nobel laureate Harry Kroto even weighed in with a video message to the university senate. The plan for a chemical-biology department has

now been dropped, says Lawless, but chemistry still faces an uncertain future.

Smith has also been criticized by the UK Parliament's Science and Technology Committee for not consulting with members of Sussex's chemistry department before unveiling his plan. The committee will publish a report on 4 May calling for greater efforts to sustain university chemistry departments.

Smith, a professor of economics, told the committee at a hearing on 27 March that he did not accept that "a serious science university must have a chemistry department". "I

absolutely stand by that," he told *Nature* last week, insisting that universities must evolve with student demand. But opponents of the closure claim that the department has been allowed to go to seed, with no effort made to retain or replace lost staff.

So the campaigning continues, with chemistry students at the university staging protest meetings and collecting more than 1,000 petition signatures and messages of support. The drive has been boosted by offers from Sussex's life-sciences departments to temporarily forgo making new appointments so that the cash can instead be funnelled into chemistry.

Sussex would be the latest in a string of department closures that have left UK chemists

"Government funding ignores how expensive chemistry is to teach compared with other subjects."

Messages of rebuke

"You are making a horrible mistake. To close the department will be an enormous loss to Sussex, to Britain and to science around the world."

Alfred Bader, Austrian founder of the Aldrich Chemical Company, now Sigma-Aldrich

"The neglect of chemistry at Sussex over recent years is nothing short of a disgrace."

Letter signed by 20 members of the **University of Oxford's chemistry department**

"As a Frenchman, I should perhaps be goading you into pursuing this unwise decision. A shortage of good British chemists will be a godsend for our young graduates."

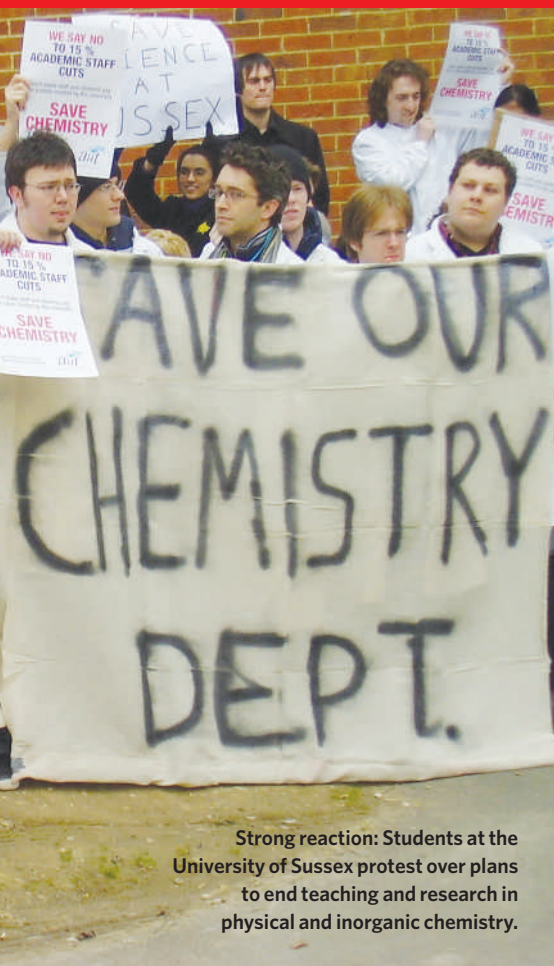
Samir Zard, chairman of chemistry at the École Polytechnique in Palaiseau, France



increasingly concerned for the future of their field (see 'Recent closures'). But with a rating of 5 (the second highest score possible) in the government's latest research assessment exercise, it is the highest-profile case so far.

"Every chemistry department is operating at a deficit," says Richard Pike, chief executive of the London-based Royal Society of Chemistry. He laments the fact that government funding ignores how expensive chemistry is to teach compared with other subjects. Laboratory space, fume hoods and equipment all send costs soaring, but chemistry departments are given just 1.7 times as much money per student as history or English departments, compared with 4 times as much for medical students. Because the subject underpins so many industries, not least the pharmaceutical business, Pike argues that investing in it is essential for the country's economic future.

"Closing chemistry departments is ultimately suicide for the country," agrees Keith Smith, head of chemistry at the University of Wales Swansea, which will teach its last group of undergraduates next year. Fresh crops of graduates are clearly essential to maintaining chemistry research in Britain, but Smith also points out that small departments such as Swansea's produce a relatively large proportion of chemistry teachers. Without them, he fears that secondary-school students will increasingly be taught by teachers without a chemistry degree, who may lack the enthusiasm to



Strong reaction: Students at the University of Sussex protest over plans to end teaching and research in physical and inorganic chemistry.

P. CECIL

Recent closures

2005: University of Exeter merges reduced chemistry department into biological sciences

2005: University of Dundee closes divisions of physical and inorganic chemistry

2005: Queen Mary, University of London, merges chemistry with biology

2004: University of Wales Swansea stops taking in new chemistry undergraduates

2004: King's College London closes chemistry department

inspire future generations to pursue the subject.

It is hard to imagine the same thing happening in the United States, says Catherine Drennan, a chemist at the Massachusetts Institute of Technology (MIT) in Cambridge. Chemistry departments are considered indispensable by US universities, she says, because the basic science they teach is relevant to so many other disciplines. Every student at MIT learns some chemistry, for example, regardless of their subject area. "Chemistry is one of the absolute core sciences," she says.

GETTY



WRINKLED CELL NUCLEI MAY MAKE US AGE

Blocking an aberrant protein could keep cells pert and young.

www.nature.com/news

"You really need those fundamentals."

Germany experienced a similar chemistry crisis in the 1990s, says Kurt Begitt, director of education for the German Chemical Society. This resulted in several department closures and mergers. But a renewed drive to attract students to the subject has been successful. "Departments are full," he says, adding that packed schools are difficult to close.

Back in Britain, Lawless says his fund-raising effort has already won interest from two companies, but he is reluctant to name them at such a "delicate stage of negotiations". But their offers require Smith's support, and there are yet more hurdles to face in the coming weeks, as Lawless's plan faces a plethora of university committees. If it gets through them all, a final decision on the department's future is due from the university's council on 15 May.

Frustratingly for Lawless, Smith says the meetings may merely conclude that more discussion is needed. And if they reject the plan outright, "I've no idea what we'll do", sighs Lawless. "I suppose we'll have to look at closing the department."

Mark Peplow

Regulator seeks cause of death in gene-therapy trial

German researchers are attempting to discover the cause of death of a 28-year-old man who had undergone gene transfer to treat an immunodeficiency disorder called chronic granulomatous disease.

The patient died of blood poisoning and organ failure on 10 April. Researchers on the trial say that his death was probably not directly caused by the gene insertion, unlike fatalities on gene-therapy trials for another immunodeficiency disease, SCID.

Klaus Cichutek is head of gene therapy at the Paul Ehrlich Institute, the German regulatory authority in Langen. He says that the inserted gene, which was designed to protect the man against infection, may simply have stopped working.

Top US scientists publicly condemn Guantanamo

Prominent scientists in the United States have taken the unusual step of protesting about an issue unrelated to their work: the treatment of prisoners at the US naval base at Guantanamo Bay in Cuba and at other prison camps run by the US government.

In a letter that was published on 30 April in *The New York Times*, physicist Leonard Susskind of Stanford University in California and 18 others condemn the treatment of prisoners and “disdain for international law” at the sites. “Although this is not a scientific issue in the usual sense, we feel that to ignore it would be to abdicate our responsibility to the truth,” the letter states.

Signatories to the letter include the physicists Edward Witten of the Institute for Advanced Study in Princeton, New Jersey, and Frank Wilczek of the Massachusetts Institute of Technology. “Our country has become anaesthetized, and everyone, including scientists, should speak out more,” says Susskind.



GETTY IMAGES

The treatment of prisoners at Guantanamo Bay has been criticized by US scientists.

Cloud satellites finally take to the skies

It's been a long road, but two satellites that will study how clouds affect our climate are finally in orbit around the Earth.

Clouds absorb and reflect the Sun's radiation, helping to regulate Earth's climate in a way that depends on their altitude, density and colour. But these effects are poorly understood, which introduces uncertainties into climate models.

CloudSat and CALIPSO launched on 28 April and will fly in formation to take measurements of the same part of the sky. The satellites will use radar and its visible-light equivalent, lidar, to reveal how water droplets and aerosol particles are distributed inside clouds.

The launch was initially delayed from



November 2005 when Boeing machinists went on strike. Last week saw a series of further holds-ups, on one occasion due to very thing the satellites are due to study: heavy cloud.

Medical journals told to advertise goods not drugs

Medical journals should stop accepting adverts from drug companies, say Adriane Fugh-Berman of Georgetown University School of Medicine in Washington DC and her colleagues in *PLoS Medicine*. Accepting such income, says Fugh-Berman, puts journals at risk of losing their objectivity and of implicitly endorsing certain drugs.

It is no surprise that drug companies want to advertise in journals that doctors read. But the researchers say that journals encourage the submission of drug ads over those for other kinds of product. They scoured the advertising policies of nine major medical journals and found that four accepted only ads related to medicine.

Although journals say that they rely on the drug companies for advertising revenue, researchers argue that the threat of lost income can influence editorial decisions. They suggest that journals could make as much money from ads that sell luxury goods to doctors.

Planetary scientists vote for research not missions

US planetary scientists have a request for NASA: if you must cut the budget, at least try to leave research grants alone.

That's the clear message from a poll of 1,024 scientists who answered a survey conducted by the Planetary Science Institute (PSI), based in Tucson, Arizona, and three other institutions. PSI director Mark Sykes says that about half of all US

planetary scientists responded, which "speaks to a lot of frustration in the field right now". NASA wants to scale back its Solar System research from \$351 million to \$273 million by 2007. Congress is reviewing the proposed cuts.

Almost 90% of respondents ranked research and analysis grants as their first or second priority for funding, ahead of spacecraft missions. The scientists wouldn't ditch the expensive missions altogether, however — most would choose to fly fewer smaller missions to pay for saving some larger projects.

Scrapping funding review will cost more, UK warned

Plans to scrap Britain's Research Assessment Exercise (RAE) could end up costing more than they save, warns a report from the Higher Education Policy Institute (HEPI).

A proposal to end the RAE, which distributes more than £1 billion (US\$1.9 billion) every year to higher-education institutions for core costs such as equipment and salaries, was made this March in the UK budget. The RAE uses peer review, but the Treasury argued that a metrics-based approach, under which funding is allocated on the basis of external research income, would be quicker and cheaper to operate.

HEPI says that researchers would apply for more grants, driving up administration costs and potentially costing £100 million extra a year, even after savings from scrapping the RAE are taken into account. The report also warns that some universities could see their funding fall by a third under the system.

Standing firm: Elias Zerhouni, the director of the National Institutes of Health, has come under fire for encouraging applied research.



FACING THE OPPOSITION

Elias Zerhouni has one of the biggest jobs in biomedical research — running the massive US National Institutes of Health. But is he leading the agency up the right path? **Erika Check** examines his tenure.

On 6 April, Elias Zerhouni spent the morning on Capitol Hill, telling Congress to invest more money in the National Institutes of Health (NIH) — for the good of the country and its research enterprise. But even as he was making his case, scientists across the continent were e-mailing each other an angry editorial that blames Zerhouni for leading the NIH in the wrong direction.

Published in the *Journal of Clinical Investigation*, and penned by the journal's editor-in-chief Andrew Marks (A. R. Marks *J. Clin. Invest.* **116**, 844; 2006), the piece hits Zerhouni in a sensitive spot. "The current state of the NIH," writes Marks, "prompts me to say to its director, Dr Elias Zerhouni: 'Obviously you are not a scientist.'"

Marks says he meant no personal disrespect for Zerhouni, who is a radiologist by training. "There's a real gap between the people that run the NIH and the people that depend on it for their survival to do research," says Marks, a physiologist at Columbia University in New York. He says he has received a steady stream

of e-mails from other researchers, praising him for writing the piece. "The tone of the editorial reflects a level of frustration that is not unique to the author," says one senior biomedical lobbyist in Washington DC, who asked not to be named.

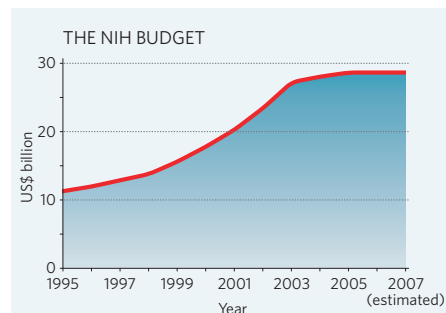
Zerhouni's many high-profile supporters say he is steering the NIH on the right course at a crucial time in the agency's history; after an unprecedented doubling of funds between 1998 and 2003, the NIH's budget has recently flattened and may even drop in the coming year (see graphic). But Zerhouni's vision of

streamlining the agency's management and sharpening its focus on clinical, interdisciplinary and technology-driven research has upset others, who say this is distracting the agency from what it has always done well: fund basic research.

Zerhouni cuts a sharp figure in Washington; he is charming and well-respected. But the former competitive swimmer is also unafraid to tussle with his critics. The changes he is making, he argues, are good for the NIH and for science. "You have to stand up for the integrity of the agency, no matter how popular or unpopular that is," he says.

Hard act to follow

The job of the NIH director is never easy, but Zerhouni inherited a particularly tough challenge in following Nobel laureate Harold Varmus, who headed the agency for six years. Previous directors were noted for either their administrative skills or their scientific achievements, but Varmus was a rare combination of world-class scientist and effective political operator. He left in 1999, and the Senate confirmed



FROM BENCH TO BEDSIDE

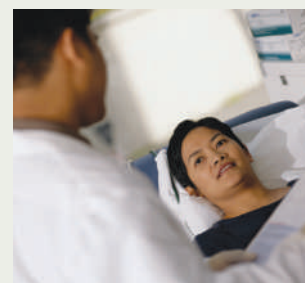
The promotion of translational research — basic-research discoveries that find their way directly into new treatments — may be the hardest pill to swallow for many of the scientists funded by the National Institutes of Health (NIH). But this issue is key for the agency's director Elias Zerhouni. "I'm very concerned that we have a disconnect

between basic science and translational science," he says.

To try to spur change, Zerhouni allotted 40% of Roadmap dollars to projects that aim to change the way translational science is done, and created a programme called the Institutional Clinical and Translational Science Award. The agency intends to give out

60 of these institutional awards, worth a total of \$500 million, by 2012.

But Zerhouni's commitment to translational research goes beyond the Roadmap itself: it permeates his entire leadership of the agency. "He chose to articulate a slightly different vision for the NIH, connecting it more intimately to health and



medicine," says one biomedical research lobbyist. "How that plays out will be his legacy." **E.C.**

President George W. Bush's nomination of Zerhouni to replace him on 2 May 2002.

Some researchers were sceptical from the start. When he took the job, Zerhouni looked more like a successful chief executive than a traditional bench scientist. A radiologist, who did research on advanced imaging techniques and patented eight inventions, his many roles at the Johns Hopkins School of Medicine in Baltimore, Maryland, included being an administrator for six years. Helping run one of the country's premier translational science institutions had, he felt, prepared him for modernizing the NIH and its mission.

When Zerhouni took the helm of the NIH, agency veteran Ruth Kirschstein had been serving as interim director, but of the 27 institutes and centres that comprise the NIH, 6 were without a permanent leader. And there was no clear plan for how the NIH would take advantage of its unprecedented rise in funds. "You couldn't help but think that you needed to create a vision for the NIH," Zerhouni says.

A year and a half after arriving at the agency, Zerhouni unveiled his signature 'Roadmap' initiative. This was designed to address a frequent criticism of the NIH: that it doesn't do a good job funding cross-disciplinary, risky and large-scale science.

More for the money

The Roadmap encourages translational research, or science that takes findings from the lab to the clinic (see 'From bench to bedside'). And it has created funding mechanisms

to promote innovation, such as the lucrative Pioneer Award for high-risk research. The Roadmap also funds team resources, such as chemical databases that can be searched for drug candidates.

At first, the Roadmap — projected to cost \$2.1 billion over five years — drew widespread praise as an attempt to reform some of the NIH's old ways of doing business. But soon after it was unveiled, Congress began passing budget increases that were far below what the NIH had been used to.

In his editorial, Marks taps into researchers' frustrations with the Roadmap. Some thought it should not have been started simultaneously with the flattening of the budget, and others felt it diverted funds away from the individual-focused, investigator grants called R01s on which basic-science researchers depend. "The NIH Roadmap is not a roadmap at all, but a yellow-brick road," writes Marks. "It looks like it will lead us back to Kansas, but the way is really fraught with danger, the end of the road is not really where we want to go, and it is all just a fantasy."

Marks told *Nature* that he feels a basic scientist would not have unveiled the Roadmap at a time when money was tight. "If the director of the NIH had a deep understanding of what it's like to be in the trenches supporting your labs with R01 grants, the timing would have been different," he suggests. "I thought that was perhaps insensitive."

Others think Marks himself was insensitive for penning the editorial. "The insulting nature of the whole piece was destructive to what we all need to do in the community," says Varmus, now president of the Memorial Sloan-Kettering Cancer Center in New York. An NIH advocate who did not want to be named, but who has worked on Capitol Hill on and off for years, put it more strongly: "He handed the people who want to hurt us a bat."

Poor connections

The editorial seems especially vitriolic because it echoes an unstated feeling in the community that Zerhouni is not 'one of the group'. Some say there is little he could do to overcome this bias. "Zerhouni doesn't have the lab-rat appearance, and there have been elements in the community that expect the NIH director to be exactly in their image," says Varmus.

Zerhouni served as a consultant to the White House when Ronald Reagan was president. That and his nomination by Bush — whose administration has been accused of ignoring science at the expense of ideology — has not endeared him to many scientists. "I think that in a more open administration, Elias would have been a very good leader," says Jerry Keusch, associate dean for global health at Boston University in Massachusetts and former director of the Fogarty International Center at the NIH. For his part, Zerhouni says: "There is a little bit of a political overture in the comments I hear that suggests scientists are opposed to this administration. I'm pretty independent as a director, and that's unfair."

But others say Zerhouni could have strengthened his position by reaching out to researchers, both inside and outside the agency. For instance, embarrassing revelations about financial conflicts of interest in newspaper reports in December 2003 and in subsequent congressional hearings (see 'Congressional conflicts'), caught Zerhouni off-guard. "If he was the kind of leader who could have understood or connected better with the institute directors and bench scientists, it would have served him much better in that storm," says one researcher, who asked to remain anonymous. "He got no

"The NIH Roadmap is not a roadmap at all, but a yellow-brick road." — Andrew Marks

MONEY WORRIES

Scientists and advocates of the National Institutes of Health (NIH) are anxious about the funding outlook for biomedical research. After the agency's budget rose from \$13.6 billion in 1998 to \$27.3 billion in 2003, there began what some call the 'undoubling' of its budget.

The president's budget request to Congress this year included the first proposed cut in NIH



spending in 36 years.

Some argue that young investigators could be especially hard-hit, and that

the investment in the doubling could rapidly be eroded. Biomedical advocates are scrambling to figure out how to reverse the trend, or at least keep the cuts from growing deeper.

"We're losing buying power every year," says Harold Varmus, president of the Memorial Sloan-Kettering Cancer Institute in New York and a former director of the NIH. **E.C.**

Low morale: after a doubling of its budget, the National Institutes of Health is now facing tougher times.



backup, and he did contribute to that.”

Zerhouni attributes much of the discontent to the flattened budget, and notes that competition for NIH grants has accelerated dramatically in recent years. From 1999 to 2003, the agency attracted 5,300 new grant applicants. Then, the agency gained an additional 5,000 new applicants just in 2004 and 2005 combined. Zerhouni calls it “a perfect storm that is hitting the NIH”.

Some say it was clear that the agency couldn't keep getting double-digit increases for ever, and that the NIH leadership should have devised a better plan earlier on for dealing with the end of the budget hike. And even though Zerhouni arrived towards the end of the budget doubling, his critics say he hasn't done enough to soften the blow. “I think he's dealt with it through prayer,” says Arthur Caplan, a bioethicist at the University of Pennsylvania in Philadelphia. “It hasn't been addressed head-on.”

Ringing the changes

But an analysis by Howard Garrison, director of public affairs for the Federation of American Societies for Experimental Biology, suggests that the total proportion of funding for R01 grants, seen as the mainstay of traditional investigator-initiated research, is about the same as it was before the doubling began. “What has happened is that the awards have gotten bigger and the competition has gotten fiercer; expectations have risen, and the number of new grants has not gone up,” Garrison says. Zerhouni says this means that the Roadmap is not killing off investigator-initiated research, as critics have charged.

Many scientists defend Zerhouni's decision to launch the Roadmap when he did. They argue that it was an astute political move at a time when the NIH needed a new vision to

move forward from budgetary and conflict-of-interest woes. Some researchers think the Roadmap helped win congressional support by showing that the agency had a bold vision that would guide new spending.

CONGRESSIONAL CONFLICTS

In January 2004, Congress began investigating conflicts of interest for the National Institutes of Health (NIH) investigators who consulted for private companies.

The NIH director Elias Zerhouni initially tried to take a moderate approach. But Congressional outrage grew in 2004; in one particularly embarrassing moment, Zerhouni was confronted with cancelled cheques showing that NIH scientists had accepted consulting payments that they hadn't revealed to their bosses. In February 2005, Zerhouni announced stringent restrictions — later tempered — for NIH employees on consulting and investing in biomedical arenas.

Many employees felt stung by the investigations, and are dismayed at the increasing paperwork and restrictions imposed on the agency. Some blame Zerhouni. “It's hard to figure out what are his decisions and what are the administration's decisions,” says one NIH employee, adding, “Internally, morale is a problem.”

E.C.



“I share with Elias the feeling that the Roadmap stimulates Congress to reinvest, to get over their feeling that they've doubled the budget and nothing more is needed,” says Gerald Fischbach, who headed the NIH's National Institute of Neurological Disorders and Stroke from 1998 to 2001, and who is now vice-president for health and biomedical sciences at Columbia University.

Zerhouni argues that the research community has long called for the NIH to become a more flexible, forward-thinking enterprise that is in touch with the emerging need for technology-driven, interdisciplinary work. A 2003 report from the Institute of Medicine, ‘Enhancing the vitality of the NIH’, explicitly called for such an approach. “I've not yet heard a scientific criticism of the Roadmap,” says Zerhouni. “I've only heard: ‘It's taking away money from my programme or my institute.’”

With no term limit, Zerhouni could potentially stay in his post for as long as it takes to fully implement the Roadmap. As stormy as his first four years have been, it's not likely to get much easier, say observers such as former Congressman Jim Greenwood, who investigated conflicts of interest at the NIH. Greenwood, now president of the Biotechnology Industry Organization, based in Washington DC, gives Zerhouni high marks for starting the NIH down a new path. But he says it will take time for the agency and its constituents to come along.

“This is about turning around an aircraft carrier,” he says. “You can't measure the radius of the turn at first. But you don't make the turn unless you take the wheel and start spinning it — and that's what he's done.”

Erika Check is Nature's biomedical correspondent. With additional reporting by Colin Macilwain.



The NIH should make the transition to a research career smoother for young scientists, says Tshaka Cunningham.

IT'S A POSTDOC'S LIFE

Does the respected US National Institutes of Health meet the needs of young postdoc researchers? **Jacqueline Ruttimann** investigates.

Place yourself, for a moment, in the lab coat of a US postdoctoral fellow. You've just spent four to seven years working on your graduate-school dissertation, but the journey is not yet complete. At a time when many of your friends are making good money at real jobs, you have to work in another laboratory for almost the same time it took to complete grad school. Fail to receive funding and publish frequently, and your dream of one day heading your own lab will go up in smoke.

The odds are stacked against you — and the roughly 45,000 other postdocs in the country. The most recent data from the National Science Foundation show that, of those who were postdocs in 2001, less than one-quarter had moved into a tenure-track university position two years later.

Nevertheless, more than 20,000 postdocs work each year under the auspices of the US National Institutes of Health (NIH), the largest funding source for postdoc research in the country. Just over 10% of these young scientists work on 'intramural' research at the agency's main campus in Bethesda, Maryland. The rest rely on NIH funding for extramural research in labs across the country.

In many ways, the NIH is an ideal place to be a postdoc. Although its budget is being cut back now, the agency had been relatively flush with cash in recent years, and its director Elias Zerhouni has made young investigators a priority during his tenure (see page 17). Yet even at the NIH, postdocs struggle to acquire the skills needed to develop their own line of research and score a coveted tenure-track position. *Nature* spoke with four postdocs who are at key points along the journey to becoming a fledgling scientist.

Academia to industry

Sometimes the hardest thing about doing a postdoc is making the decision to become one. According to a survey by the research society Sigma Xi, the typical US postdoc makes an average salary of just \$38,000. Financial concerns can often draw young scientists off into industry or other more lucrative careers. "Those of us who decide to stick with academic science have to make a financial sacrifice that we shouldn't be making," says Tshaka Cunningham, who started a postdoc four months ago at the National Cancer Institute's vaccine branch on a project to develop an AIDS vaccine.

Even so, Cunningham is infused with the energy of a young scientist. He has lots of ideas — so many that he keeps a notebook by his bed, thinking that he might one day use some of them at a university tenure-track position or as a consultant in the biotechnology industry. For now, he hopes that the NIH can help to steer him down either path.

It's a big request for an agency reeling from a conflict-of-interest scandal, in which NIH investigators were chastized for regularly accepting large consulting fees from pharmaceutical companies and others whose products



"The problem for postdocs is uncertainty. I don't know where I'm going to be in two years."

— Stephen Sperber

S. NANDA



"There are all sorts of opportunities for the wayward scientist."

— Bernard McWatters

they evaluated. Cunningham is experiencing some of the backlash from this scandal, as everything down to his involvement with a non-profit organization he helped to found before his arrival, the National Association for Blacks in Bio, is being carefully scrutinized under new, more stringent NIH ethics rules.

Still, he'd like to see the NIH adopt a more seamless relationship with industry so that postdocs could apply for industry grants while finishing their mentor's research project. As a graduate student at the Rockefeller Institute in New York, Cunningham combined his main project with a scheme funded by a \$40,000 grant, part of which came from Merck, the New Jersey-based pharmaceutical giant. "The NIH should be an intermediary between academia and industry," he says.

That's unlikely to happen any time soon, but the NIH is focusing on smoothing relationships between academic disciplines. The agency is moving towards integrating training programmes for different disciplines, so that postdocs can more readily move between different laboratories at the various institutes, says Michael Gottesman, the NIH's deputy director for intramural research. For now, that means having bench-bound postdocs attend a weekly course called Demystifying Medicine, among others, in which they learn more about the medical research that might spring from their lab findings.

Can postdocs manage?

Career training is what many postdocs crave most, according to the recent Sigma Xi survey. Many feel they have plenty of experience and training in the laboratory, but want more structured approaches to bigger career questions. Stephen Sperber, a second-year postdoc in a molecular biology lab at the National Institute of Child Health and Human Development, says that although he enjoys his work, the future is a bit worrisome. "As comfortable as I am now, this is not a job," he says. "That's the problem for postdocs: a lot of uncertainty. I don't know where I'm going to be in two years."

Like most, he would like to run his own laboratory one day. But that will take some luck. Regardless of where they train, the average age at which a postdoc gets awarded his or her first independent research grant has been creeping up over time — from age 37 in 1980, to 42 in

2002. Only 6% of the much-desired 'R01' research grants go to new investigators.

"That transition from postdoc to primary investigator is a very difficult one," admits Gottesman. One way the NIH is trying to help is through the new Pathway to Independence award, in which researchers in the middle of their postdocs can apply for a bridge award of up to \$1 million — to fund the rest of their postdoc years and the beginning of their independent research. The award, says Gottesman, is "a morale builder".

Postdocs generally agree, but say launching a career in science takes more than just money. Many young postdocs often complain that they need management skills, training which usually takes a lower priority than getting a particular experiment done or a research paper written up. For them, the NIH has set up a course called Making the Right Moves: A Practical Guide to Scientific Management for Postdocs and New Faculty, says Norka Ruiz Bravo, deputy director for extramural research. The problem is, many postdocs are often not aware of such training opportunities, because they are too isolated in their particular institute or simply have not been able to find the information they need.

Alternative paths

There are also those who decide that bench life is not for them. Bernard McWatters has just finished a postdoc on a project linking polio vaccines to cancer. An affable young man with twinkling eyes and a mouth usually curled up in a smile, McWatters was extremely focused when it came to looking for a job. He knew he



"You're not sure whether you want to go back home and you're not sure whether you want to stay."

— Alessandro Vichi

wanted to work at the US Food and Drug Administration (FDA) and achieved it by doing a hybrid postdoc between the FDA and the NIH, in which he spent half of his time at the bench and the other half doing regulatory work. He will soon start a regulatory-affairs job at the FDA's Centers for Biological Evaluation and Research. "There are all sorts of opportunities for the wayward scientist looking for a job," he says.

McWatters ought to know. He has served as chairman of a career development subcommittee of a larger group set up to improve communication between postdocs and the NIH community. McWatters helped to sponsor



"It's a remarkable time of your life. A time when you focus on science."

— Norka Ruiz Bravo

monthly seminars on alternative careers, including such fields as patent law, science policy, medical communication and science consulting. He advises postdocs not to focus on just one career path, and not to wait until the end of their postdoc to start looking for jobs.

The job hunt brings other complications for international postdocs. More than two-thirds of the NIH's intramural postdocs arrive from other countries, often on a visa that ties them to a particular mentor's lab. Alessandro Vichi, who came from Italy to work as a postdoc for five years at the NIH, says that the academic pressures are similar for US and international postdocs. But those visiting from abroad have to worry about visa issues, particularly since regulations were tightened after the terrorist attacks of 11 September 2001. "You're waiting without making a plan," says Vichi, who is now a research fellow at the National Heart, Lung, and Blood Institute. "You're not sure whether you want to go back home and you're not sure whether you want to stay."

Regardless of whether foreigners decide to stay or leave the country after their postdoc, Gottesman argues that all receive sufficient training to get jobs. Those who stay, he says, are often successful in finding work — even if they are in the biotech industry or government rather than an elusive tenure-track position. "Unemployment among people with postdoc experience is very low," he says.

But even NIH officials admit that there is little evidence to support this reported success. The NIH has no formalized tracking system to follow where postdocs end up, which makes it difficult to monitor the success of their training. Bravo says that a tracking system is in the works, but is proving difficult to design as each of the agency's 27 separate institutes and centres have postdocs supported on different types of fellowships.

Despite all the trials and tribulations that come with being a postdoc, it is often considered one of the most productive periods of a scientist's career. "It's a remarkable time of your life when you're unencumbered with the responsibilities that go with becoming an independent investigator," says Bravo, who fondly remembers her days as a postdoc. "It's a time when you focus on science." ■

Jacqueline Ruttimann is an intern in *Nature's* Washington office.

NIH

A. VICH

BUSINESS

The quiet rise of the clinical contractor

More and more drug trials are being farmed out to contractors, but who's benefiting from the change? Meredith Wadman investigates.

When a big US drug company needed to get a late-stage breast-cancer trial launched quickly some years ago, it turned to PRA International, one of a dozen major contract-research organizations (CROs) that now permeate the worldwide drug industry.

PRA is headquartered in a nondescript office building outside Washington, DC. But its reach is global, encompassing 23 countries on six continents. PRA quickly put its connections to use, recruiting 800 trial subjects in Poland, Russia and Mexico at ten times the rate that had been achieved in US hospitals, clinics and doctors' surgeries. "I can't believe the enrollment... Every time I check, I am stunned," a delighted drug-company official wrote to the contractor.

Not everyone is so enthusiastic about the rise of the contractors, who are managing an increasing proportion of clinical trials. Their role was spotlighted in March, when a trial at Northwick Park Hospital in London, managed by Parexel, a Massachusetts-based contractor, went badly wrong (see *Nature* 440, 388–389; 2006). Critics claim that the involvement of contractors adds an obscuring layer to a process that should be transparent to the public — and worry that in their quest to deliver data fast, contractors may cut corners.

"Their *raison d'être* is to speed things up. And speed and safety are anathema to each other," says Vera Hassner Sharav, president of the Alliance for Human Research Protection, a research watchdog group based in New York.

The total value of contract research in the drug industry has grown spectacularly from US\$1.6 billion in 1993 to about \$15 billion this year (see graph). And New York investment bank Goldman Sachs predicts it will reach \$26 billion by the end of the decade.

PRA is typical of this pattern. Since 1999, it has carried out about 2,800 clinical trials, including key studies of blockbuster drugs such as Genentech's Avastin for cancer. In that time, its revenues have tripled, to almost \$300 million in 2005, and its staff has grown from

1,000 to 2,400, with 400 more to join this year. Goldman Sachs has just upgraded the company's share rating to "outperform".

The key to this success is a shift in the behaviour of the pharmaceutical industry, which 20 years ago ran virtually all its own drug development. Today, by most estimates, drug-makers are contracting out around 20% of the work involved in taking a drug from discovery to the market.

"It has become obvious to most pharmaceutical companies that outsourcing is a good way to keep costs down," says David Dockhorn,

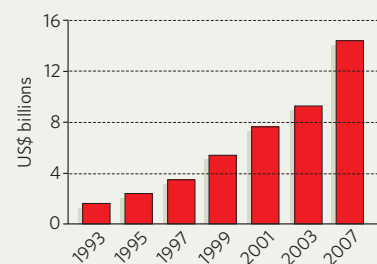
"It has become obvious to companies that outsourcing is a good way to keep costs down."

PRA's clinical vice-president. He adds that contractors offer drug-makers a valuable commodity — experience. "In a single year we will see five to ten compounds in a therapeutic area. Scientists in a large pharmaceutical company may see one or two in their entire careers."

As well as preparing and conducting clinical trials, PRA analyses data collected by companies once their drugs are on the market. Other contractors also offer additional services, ranging from animal testing to specimen analysis.

The rise of the contractors has helped the pharmaceutical industry trim fixed costs and protect themselves from the booms and busts caused by erratic drug pipelines. "What pharmaceutical companies have done is let the CRO industry weather those fluctuations," says David Orloff, medical director of Medpace, a clinical contractor based in Cincinnati, Ohio. "That way they don't saddle themselves

CONTRACT RESEARCH ORGANIZATIONS: GLOBAL REVENUE



with the burden of having to manage their overhead through times of plenty and times of relative scarcity."

The contractors also bring global access to investigators and patients. As experimental drugs become more specialized and narrowly targeted, appropriate trial populations have become tougher to recruit, especially in North America, Europe and Japan. And running a trial in the developing world can cost a third of what it does in a wealthy country. According to the Center for the Study of Drug Development (CSDD) at Tufts University, Massachusetts, the fraction of US clinical trials conducted abroad has risen from 6% in 1991 to almost one-third last year.

The maturing of the biotechnology industry, whose smaller companies do not usually have their own clinical trials infrastructure, has also generated new customers for the contractors. PRA, for example, draws half its business from biotech companies.

There is conflicting data, however, on whether contractors actually conduct clinical trials more efficiently. A study led by Ken Getz of the CSDD and published in January this year found that submissions to regulators from trials in which contractors were heavily involved were made on average 30 days earlier than those from other trials. But another investigation, completed last year by Chicago-based management consulting firm KMR, reported that the more-efficient drug companies — those that spent less to get trials done — were no more likely to use contractors than were their less efficient competitors.

The Tufts study also found that, since 2001, drug-makers have boosted their spending on CROs by 15% annually, whereas their overall



Clinical contractors have been put under the spotlight by a recent drug trial in London.

spending on drug development grew by 11%. "The factors driving growth now are not going to change," says Getz.

Parexel has declined to discuss the circumstances of the London trial in detail. Six healthy volunteers became seriously ill soon after injection of a monoclonal antibody made by the tiny German firm TeGenero. They were dosed rapidly one after another — according to a protocol approved by regulators — so that all the subjects had received the drug before the first one fell ill.

Parexel did issue a statement on 5 April, the day the British Medicines and Healthcare Regulatory Agency (MHRA) released an interim report stating that nothing in the way the trial was run contributed to the volunteers' illness (see *Nature* **440**, 855–856; 2006). "The MHRA has confirmed that the trial was run according to the approved protocol," the Parexel statement said. "These findings support our internal review that best practices and policies and procedures were correctly followed."

The MHRA's interim findings "should be seen as a positive for the industry," says Alex Alvarez, an analyst at Goldman Sachs. But he cautions that the booming industry needs to watch its step. "The one area where the whole industry could fall apart is if one of these CROs decided to do something improper or skip steps simply to expedite a trial or reduce the cost of a trial." In Alvarez's judgement, industry managers are unlikely to let this happen. "They know that's probably the biggest risk to them and to the industry. That's not an area where they would be looking to cut time or costs." ■

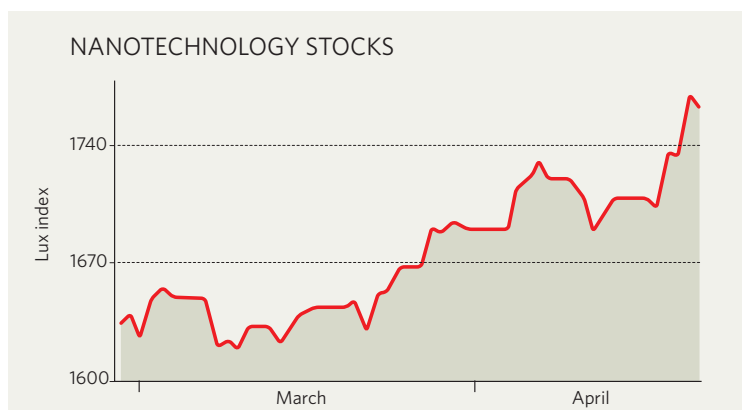
IN BRIEF

GREEN LIGHT FOR BIOGENERIC The European Commission granted its first-ever approval to a generic biological drug — for Omnitrope, a copycat version of human growth hormone, made by the Sandoz unit of Swiss drug-maker Novartis. The landmark decision was made under a 2005 law that opens a route to market for biogenerics in Europe (see *Nature* **438**, 154–155; 2005). In the United States, meanwhile, a federal judge ordered the US Food and Drug Administration to end four years of inaction on Sandoz's application to market Omnitrope in the United States.

VALLEY SUNSET Sun Microsystems founder Scott McNealy, one of the best-known business leaders in California's Silicon Valley, stepped down as chief executive of the computer company, after 22 years. McNealy, known for his hard-driving management style and his endless feuding with Microsoft's Bill Gates, will remain company chairman but be replaced as chief executive by another Sun manager, Jonathan Schwartz. Sun has struggled to cope with soft demand for its Internet servers since the dot.com boom collapsed in 2001.

VIOXX SCORECARD Merck suffered a setback in litigation over the painkiller Vioxx. A jury in south Texas ordered the company to pay \$32 million in damages to the family of Leonel Garza, a 71-year-old man who suffered a fatal heart attack after taking the painkiller for less than a month. The New Jersey company has now lost three of six Vioxx cases to reach juries, and is facing about 11,500 more lawsuits over a drug that it withdrew in 2004, after a study linked its long-term use to an increased risk of heart attacks and strokes.

MARKET WATCH



Nanotechnology stocks continued their strong run in the spring, with the Lux Research index rising about twice as fast as wider measures of the technology market over the past two months.

Strong performers during this period included Immunicon, a Pennsylvania-based company that develops magnetic particles for use in cancer diagnostics, whose shares soared from just over \$3 to touch \$5 last month after it announced the extension of a collaborative research agreement with Pfizer at the end of March.

Another stock that powered ahead since March was Arrowhead Research, a diversified Californian nanotechnology group. According to Peter Hebert, president of Lux Research, the New York consultancy that compiles the Lux index, Arrowhead

owns several small companies that may soon be ready to make initial public stock offerings.

There have been no such offerings so far this year in the United States, but Cap-XX, an Australian company that makes supercapacitors, launched successfully on London's AIM on 27 April, raising about £17 million (US\$31 million) from investors.

Hebert says that adverse reports about the safety of nanotech — notably those that followed the withdrawal of a bathroom cleaner called Magic Nano in Germany on 30 March — have had little effect on overall market sentiment. "It's a significant issue," he says, "and it's important that research is done into health and safety. But large companies are not shifting away from nanotechnology and neither are national governments."

Colin Macilwain

SOURCE: LUX RESEARCH

Computing: report leaps geographical barriers but stumbles over gender

SIR — As senior researchers in computer science, we were interested in both the report *Towards 2020 Science*, published by the Microsoft Corporation, and your related set of News Features and Commentaries (*Nature* **440**, 398–405 and 409–419; 2006). The vision of advanced computational techniques being tightly integrated with core science is an exciting and promising one, which we are glad to see being carefully explored and presented to the broader community.

We are, however, concerned that, of the 41 participants and commentators brought together by Microsoft, not one was female, with the same being true of the nine authors of the related articles in *Nature*. The report notes that the participants in the 2020 Science Group were geographically diverse, representing 12 nationalities, coming “from some of the world’s leading research institutions and companies [and]... elected for their expertise in a principal field”. Women have earned between 13% and 18% of all PhDs awarded in computer science and engineering in the United States during the past two decades. Women also work at leading research institutions, and also have expertise in the relevant fields. In most other scientific fields represented in the report, an even higher percentage of PhDs is female.

That the omission of women from the 2020 Science Group was doubtless unintentional does not lessen the negative message conveyed. The future of computing will be defined by the efforts of female as well as male computer scientists.

Martha E. Pollack

Computer Science and Engineering, University of Michigan, 2260 Hayward Street, Ann Arbor, Michigan 48109, USA

Susanne E. Hambrusch *Purdue University, USA*

Carla Schlatter *Ellis Duke University, USA*

Barbara J. Grosz *Harvard University, USA*

Jessica Hodgins *Carnegie Mellon University, USA*

Ruzena Bajcsy *University of California, Berkeley, USA*

Carla E. Brodley *Tufts University, USA*

Luigia Carlucci Aiello *Università di Roma La Sapienza, Italy*

Maria Paola Bonacina *Università degli Studi di Verona, Italy*

Lori A. Clarke *University of Massachusetts, Amherst, USA*

Julia Hirschberg *Columbia University, USA*

Manuela M. Veloso *Carnegie Mellon University, USA*

Nancy Amato *Texas A&M University, USA*

Liz Sonenberg *University of Melbourne, Australia*

Elaine Weyuker *AT&T Labs, USA*

Lori Pollock *University of Delaware, USA*

Mary Jane Irwin *Penn State University, USA*

Lin Padgham *RMIT University, Australia*

Barbara G. Ryder *Rutgers University, USA*

Tiziana Catarci *Università di Roma La Sapienza, Italy*

Kathleen F. McCoy *University of Delaware, USA*

Maria Klawe *Princeton University, USA*

Sandra Carberry *University of Delaware, USA*

Laura Dillon *Michigan State University, USA*

Kathleen McKeown *Columbia University, USA*

Mary Lou Soffa *University of Virginia, USA*

Computer ‘recycling’ builds garbage dumps overseas

SIR — Your Editorial “Steering the future of computing” (*Nature* **440**, 383; 2006) explores the future potential of the computing industry. Interesting though this is, I am concerned by the millions of tonnes of electronic waste generated by the computer industry in the United States and other developed countries each year, much of which is being shipped for recycling in developing countries such as India, China, Bangladesh and Pakistan.

Cheap labour and weak environmental standards and law enforcement in developing countries attract high-tech garbage-dumping in the name of recycling. Old computers are being dumped or burned in irrigation canals and waterways across Asia, where they are releasing toxic substances such as lead, mercury, cadmium, beryllium and brominated flame retardants that pose serious health hazards to local people and the natural environment.

The 1989 Basel Convention, restricting the transfer of hazardous waste, has been ratified by all developed countries except the United States — which, according to the environmentalist report *Exporting Harm* (see www.svtc.org/cleancc/pubs/technotrash.htm), exports 50–80% of its computer waste. Many nations, including the European Union, have gone further and ratified an amendment banning all export of hazardous waste to developing countries. Those who have not should do more towards finding solutions for the safe disposal of accumulated hazardous waste on their own territory.

G. Agoramoorthy

Department of Pharmacy, Tajen University, Yanpu, Pingtung 907, Taiwan

A logical alternative for biological computing

SIR — Roger Brent and Jehoshua Bruck, in their Commentary article “Can computers help to explain biology?” (*Nature* **440**, 416–417; 2006), draw a firm distinction between von Neumann computers — the usual computer as we know it — and biological systems. But there are many alternative models of computation. A Prolog (logic programming) computer, in particular, does not seem to exhibit several of the differences singled out.

A Prolog computation, like its biological counterpart, does not need an order of

execution. Any partial ordering of the major components, known as clauses, are determined by a dynamic succession of pattern-matching operations. Within these clauses, the execution of logic expressions is unordered: A and B is the same as B and A, and it does not matter whether we deal first with the truth of A or the truth of B (although computational constraints sometimes impose a partial ordering). A key for biological modelling would be to impose only those sequence constraints that have analogues in biological systems.

A second distinction highlighted by Brent and Bruck is that biological systems do not have a separate ‘output’ component. Again, Prolog does not conform to the norm. Often the important reason for executing a Prolog program is to find out what ‘bindings’ occur en route to a true outcome, in other words, what values are bound to what variables.

It is perhaps relevant that Stephen H. Muggleton, in his companion Commentary article “Exceeding human limits” (*Nature* **440**, 409–410; 2006), encourages the development of new formalisms within computer science that integrate mathematical logic and probability calculus.

Prolog may not be a perfect computational model for biological systems, but it exemplifies a system that could be a better fit for biological modelling.

Derek Partridge

School of Engineering, Computer Science and Mathematics, Harrison Building, University of Exeter, Exeter EX4 4QF, UK

Colossus was the first electronic digital computer

SIR — Your timeline (“Milestones in scientific computing” *Nature* **440**, 401–405; 2006) starts in 1946 with ENIAC, “widely thought of as the first electronic digital computer”. But that title should arguably be held by the British special-purpose computer Colossus (1943), used during the Second World War in the secret code-breaking centre at Bletchley Park.

Modern computing history starts even earlier, in 1941, with the completion of the first working program-controlled computer Z3 by Konrad Zuse in Berlin. Zuse used electrical relays to implement switches, whereas Colossus and ENIAC used tubes. But the nature of the switches is not essential — today’s machines use transistors, and the future may belong to optical or other types of switches.

Jürgen Schmidhuber

Dalle Molle Institute for Artificial Intelligence, Galleria 2, 6928 Manno-Lugano, Switzerland, and Institut für Informatik, TUM, Boltzmannstraße 3, D-85748 Garching bei München, Germany

BOOKS & ARTS

A Palaeozoic whodunnit

What caused a mass extinction event 250 million years ago?

Extinction: How Life on Earth Nearly Ended 250 Million Years Ago

by Douglas H. Erwin

Princeton University Press: 2006. 306 pp.

\$24.95, £15.95

Michael Benton

Science is about being wrong as much as it is about being right. Thirteen years ago, Douglas Erwin wrote in his book *The Great Paleozoic Crisis* (Columbia University Press, 1993) that the extinction event at the Permo-Triassic boundary (PTB) 250 million years ago had lasted for as long as 10 million years and had many causes: volcanic eruptions, anoxia, acid rain, climate change, massive methane release from the ocean floor, and perhaps even meteorite impact. He termed this idea the 'Murder on the Orient Express' model, after Agatha Christie's mystery of the same name, in which it turned out that all 12 passengers in the train had contributed to the murder of the American millionaire Mr Ratchett.

In his new book, *Extinction*, Erwin argues that the PTB event lasted for thousands, rather than millions, of years and that a much reduced panoply of killing agents was involved. Many of these trace back to the eruption of the Siberian Traps, 4 million cubic kilometres of flood basalt lavas that erupted for 600,000 years over a huge area of eastern Russia. In *Extinction*, Erwin documents a research programme in rapid evolution.

His earlier view was typical of the time, reflecting a partly explored topic, and the intervening decade has witnessed a remarkable advance in knowledge. In 1993, stratigraphic precision was lacking. Palaeontologists commonly identified a long-term decline in their favourite groups of organisms through the entire Late Permian, whereas we now recognize a precursor mass extinction some 10 million years before the big one at the PTB. The problem was that geologists had not established reliable dating schemes, and many fossils were tagged as little more than 'Upper Permian'. Another problem that Erwin highlighted evocatively in his 1993 book was that many of the best PTB sections lay in politically difficult areas, notably Iraq, Afghanistan, Kashmir and China.

Basic fieldwork in southern China through the 1990s has provided remarkable documentation of the Upper Permian and Lower Triassic



Looking for clues: Jin Yugan studies rocks from the Late Permian, around the time of the mass extinction.

sequences. Detailed study of the sediments and fossils has revealed the nature of the environmental changes, how the different groups of organisms fared before and after the event, and the order and timing of all this. Finally, technical improvements in radiometric dating have provided amazing levels of precision for dates on individual volcanic ash bands, with error bars as small as 0.2%; one team calculates the date of the PTB as 251.6 ± 0.6 Myr, and another as 252.6 ± 0.2 Myr, the difference depending on how the ashes are sampled.

Those who write about science should give us the grit, the sweat and the boredom. Erwin certainly gives a lively account of how the geology and palaeontology of the PTB event have been disentangled, and he has been a part of many of the advances. He takes us to the Guadalupe Mountains of Texas, Meishan in southern China, and the Karoo in South Africa. Erwin also introduces the scientific methods and debates in the field and laboratory. He writes of the enthusiasm and commitment of the geologists and palaeontologists in many lands, but portrays little of the hard work involved. One key paper on the palaeontology of the event, for example, was by Jin Yugan and colleagues (*Science* **289**, 432–436; 2000), who showed that 94% of marine species died out precisely at the PTB in the Meishan section in south China, and that further steps of extinction continued for 200,000 years or more after that first crisis. This

apparently simple result was based on hundreds of thousands of specimens of 333 species, documented by taking rock samples every few centimetres through 80 metres of rock section, and sorting them back in the lab. Misquoting Thomas Edison, scientific advances are 1% inspiration and 99% perspiration, and I feel we need more of the perspiration in this book.

Erwin is a palaeontologist, and that aspect of the story is well told. After all, the essence of a mass extinction is its biological impact, and yet extinction research is often hijacked by the geological factors that may have caused such a devastating loss of life. It's probably easier to pinpoint a massive volcanic eruption or an asteroid impact crater than to document the decline of ecosystems over time. Biology is complex, and it is no fault of the palaeontologists that they often fail to nail the answers to questions such as what caused the extinction, whether it was ecologically selective, and how life recovered afterwards. This was, after all, the nearest life ever came to annihilation — perhaps only 5% of species survived. And if the causes were Earth-bound — massive volcanic eruptions leading to acid rain, atmospheric heating and anoxia — then they hold more profound significance for many of our current concerns than if the event had been triggered by the *deus ex machina* of an asteroid impact. ■ Michael Benton is in the Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK.

The source of the problem

When Rivers Run Dry: Water — The Defining Crisis of the Twenty-First Century by Fred Pearce

Eden Books/Beacon Press: 2006. 368 pp.
£18.99/\$26.95

Ganesh Pangare

A global water shortage is set to be the major crisis of the twenty-first century, so of course a vast amount has been written about it, both in the popular media and in academic books and journals. What is unique about *When Rivers Run Dry* by journalist Fred Pearce is the style and presentation of the facts and events that explain the causes and scale of the problem. Pearce's anecdotal style and the personal accounts of the people he met during his travels make it highly readable.

Pearce takes the reader on a journey around the world exploring the worst water-related disasters. They have occurred not only in the developing world, but also in countries that on the face of it seem to have their water problems under control. How many of us are aware that the once mighty Rio Grande has been reduced to a trickle near the US–Mexican border? Or that Libya's network of pipes known as the Great Manmade River can carry 6,500,000 m³ of water every year across 600 miles of desert? Or that Saudi Arabia grew all its wheat, and now grows its alfalfa, using very old, irreplaceable 'fossil water' from ancient aquifers? Or that the groundwater used by China, India and Pakistan is estimated to be more than half of the total used by the whole world? Or that in California, 15% more water is being pumped out than is being replenished?

If the author's intention is to shock the reader with these facts, he succeeds. However, the book tends to look more at the problems and less at the proposed solutions. Also, some of the stories — such as that of the Indian farmer who uses groundwater to grow alfalfa to feed his cattle and is part of a flourishing dairy industry — need to be viewed not only in the context of the water problem, but within the context of the larger socio-economic problems in the region. Nevertheless, despite focusing solely on water, the book contains much to appreciate.

Its value lies in the way it makes such a serious subject interesting and informative both for lay readers and water professionals. It looks at the complete spectrum of the social, political and economic complexity of water, covering issues related to surface water (rivers, lakes and wetlands), groundwater exploitation and quality, conflicts related to water, trans-boundary waters, floods and droughts. Crucially, many of these are viewed through the eyes of local people who are directly affected by these issues.

Interspersed in the text are large amounts



AP PHOTO/A CRITICA/E. QUEIROZ

All dried out: a boat tries to negotiate what remains of a section of the Amazon River in Brazil.

of data, but they are explained in a way that makes them easy to grasp. For example, figures for water consumption and use are defined in the context of our daily routine and activities, such as the amount of water required to grow various foods, where this water comes from, and whose water we are taking when we consume these foods. Comparisons between the way countries use and consume water are made in a similar manner. These facts are also used to explain the politics and economics of 'blue' water (found on the surface) and 'green' water (stored as soil moisture and hence available to plants), as well as the import and export of virtual water, concepts that previously remained in the domain of water professionals and academics.

Rivers are sacred to religions, symbols of

nationhood, and givers of life. We need water to survive, today and in the future. The desire to meet today's needs has always influenced the way we deal with water and forced us to overcome our desire to save some for future use. But this must change or there will soon be almost none left for tomorrow, and this is the book's central message. We need to save our rivers. We need to find ways of "storing water without wrecking the environment, of restoring water to rivers, and refilling lakes and wetlands", and of sharing waters rather than fighting over them. "It requires us to go with the flow. And to do so before the rivers finally run dry."

Ganesh Pangare is director of the World Water Institute, 6 Pentium Classic Apartments, NDA-Pashan Road, Bavdhan, Pune 411021, India.

A martian mystery

The Rock from Mars: A Detective Story on Two Planets

by Kathy Sawyer

Random House: 2006. 416 pp. \$25.95

John F. Kerridge

Discovering life beyond Earth will change forever our view of the Universe and our place within it. So when a team led by David McKay of NASA's Johnson Space Center announced in August 1996 that they had found evidence for ancient life in a martian meteorite (D. S. McKay *et al. Science* **273**, 924–930; 1996), the world sat up and took notice. How that announcement came to pass and what happened afterwards form the subject matter of *The Rock From Mars*.

Kathy Sawyer, a journalist formerly with *The Washington Post*, has produced a model of science writing for the general public. She gets the science right (with an occasional bobble), she reports with commendable balance on the intense controversy generated by McKay and colleagues' announcement, and her lucid writing should prove largely comprehensible to non-scientists. She accurately depicts the day-to-day life of the scientists involved and brings out how their observations and ideas are processed by the scientific community as it gropes its way towards the truth.

McKay's team got one thing dead right: if you want the scientific community to believe you've found evidence for extraterrestrial life,

you should provide several independent lines of evidence, something Sawyer terms a 'holistic' approach. The lines of evidence they use are termed 'biosignatures'. The critical feature of a biosignature is not that it can be produced by biological activity, but that it cannot be produced non-biologically. Evaluation of McKay and colleagues' claim therefore rested on whether their lines of evidence were really biosignatures, or whether they could have been produced non-biologically.

Their observations of meteorite ALH84001 (which is firmly established as martian even though it was found in Antarctica) led McKay and colleagues to advance four potential biosignatures. Two of them — the presence of organic matter and ultra-small fossil-like forms — were quickly rebuffed. The organic matter, although partly of apparently martian origin (itself a notable achievement by the team), consisted of polycyclic aromatic hydrocarbons, which are not plausible biosignatures as they can easily be made by non-biological processes. The 'fossils' could not be reliably distinguished from textures intrinsic to their mineral hosts, which had not previously been studied at such high magnification. That left two candidate biosignatures: unusual layered globules of carbonate, and, contained within them, tiny magnetite crystals that closely resembled those in magnetotactic bacteria on Earth.

Towards the end of chapter 13, Sawyer cites several studies that "soon challenged the biological scenario anew". But then the chapter winds down without further reference to these studies, two of which turn out to be particularly important (they are briefly described in the notes). In the final two chapters, Sawyer shifts gears twice, once to discuss a marginally related controversy, and finally to report on some recent missions to Mars. Both chapters would have made interesting appendices to the book, but the claim made by McKay and colleagues is left hanging in limbo; the "detective story" ends not with a bang but a whimper. This is particularly unfortunate as the two studies mentioned above would have provided the bang of two smoking guns.

First, a team from the Johnson Space Center led by Gordon McKay, David's brother, succeeded in synthesizing identical layered carbonate globules in the absence of biological activity, thereby destroying the globules' status as a plausible biosignature (D. C. Golden *et al.* *Am. Mineral.* **86**, 370–375; 2001). Second, a study by David Barber and Ed Scott (*Proc. Natl Acad. Sci. USA* **99**, 6556–6561; 2002) drove the final nail into the coffin. They showed that the crystal lattices of the ultrafine magnetites within the meteorite's carbonates were topotactically related to that of the carbonate host, a point that I had made in 1999, based on findings by John Bradley and co-workers (*Meteorit. Planet. Sci.* **33**, 765–774; 1998). Note that this is a true three-dimensional relationship, as first observed synthetically by J. D. Bernal and co-workers (*Clay Mineral. Bull.* **4**, 15–30; 1959);

it should not be confused with epitaxy, a two-dimensional relationship. It means that the magnetites must have formed directly by decomposition of carbonate, and could not plausibly be bacterial in origin. Barring an extraordinary coincidence, this conclusion must apply to all the nanoscale magnetites in the carbonate.

Even though I would have liked the book to conclude with closure of this tumultuous episode, this is nonetheless an outstanding

popularization of science that deserves to be widely read, not least by those interested in the 'logic of scientific discovery'.

Finally, it is worth noting that, despite the demise of the biosignatures proposed by McKay and co-workers, the search for evidence of ancient life on Mars remains scientifically reasonable and, indeed, of fundamental importance. ■

John F. Kerridge is at 334 El Amigo Road, Del Mar, California 92014, USA.

A search for meaning

The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom

by Jonathan Haidt

Basic Books: 2005. 320 pp. £34.95, £15.50.

To be published in Britain in August by William Heinemann.

Daniel Nettle

There is a striking similarity between the advice of the ancients on how to live, and the thoughts of modern psychologists on how to have a healthy mind. The Roman emperor Marcus Aurelius' dictum that life "is but what you deem it" resonates with modern research on the importance of thinking styles in coping with stress and adversity; and the Buddha's teachings on non-attachment seem to prefigure the ideas of modern cognitive therapies. The ancients, it seems, were good psychologists and understood, in an observational and intuitive way, how the mind works.

Jonathan Haidt takes this insight seriously

in his new book *The Happiness Hypothesis*. The subtitle is a much better description of the fare on offer than the main title, as he takes the conclusions of classical philosophers and thinkers on ten enduringly important themes and considers their conclusions alongside the findings of modern psychology. The ten subject areas are: the idea that the mind is made up of several often-conflicting drives or mechanisms; the idea that how we think about the world is more important than how the world actually is; the importance of reciprocity in social life; the biases that blind us to our own shortcomings but not to shortcomings in others; the paradoxical nature of pursuing happiness; the importance of love; the strengthening power of adversity; the importance of virtue, in its broadest sense; the power of spirituality; and finally, the importance of coherence in life. Haidt draws principally from Greek and Roman philosophy, from Indian and Chinese traditions, as well as a peppering of



What does it all mean? Monty Python's *The Meaning of Life* may not have been far from the truth.

RONALD GRANT ARCHIVE

other sources. He has also read widely across various schools of modern psychology, including social psychology, the evolution of behaviour, and the emotions, an area to which he has made distinguished contributions.

This is a delightful book. I enjoyed Haidt's exploration of ancient texts, but was much more impressed by the breadth of his grasp of modern behavioural science. He does have occasional blind spots: for example, evaluating the Buddha's advice on non-attachment as if it were an empirical hypothesis seems to trivialize it, and his review of cultural group selection does not consider the alternatives. Despite this, Haidt's writing embraces spiritual

and mystical viewpoints while retaining scientific and rational coherence. This is by some margin the most intellectually substantial book to arise from the 'positive psychology' movement, which is often characterized by having too much 'positive' and not enough 'psychology'. Haidt is thoughtful and judicious, and is always concerned to relate his points back to the evidence.

What is the big idea to arise from this book? In a sense, there isn't one. Haidt quotes from the Monty Python film *The Meaning of Life*, in which the answer is given as: "Try to be nice to people, avoid eating fat, read a good book every now and then, get some walking in, and

try to live in peace and harmony with people of all creeds and nations." This spoof formulation nonetheless contains an important point: there is, in fact, no single meaning to life. However, life has a definite form, a set of recurring emotions, interactions and experiences, and even if there is no ultimate solution to the dilemmas they pose, there are ways of understanding them better and navigating them with greater wisdom and purpose. And, as Haidt has shown, the ancient sages and modern psychologists often agree on these.

Daniel Nettle is in the Department of Psychology, Brain and Behaviour, University of Newcastle, Newcastle NE1 4HH, UK.

A sense of civic beauty

A Renaissance painting from Urbino reveals the ideal city.



Martin Kemp

There is a feeling in highly developed societies with centralized systems of government that an orderly society and geometrical town planning go hand in hand. This association is not exclusive to Western societies: Inca, Indian and Chinese civic planning has exhibited similar tendencies. Nor is it exclusive to totalitarian regimes, as it has featured prominently in socialist notions of 'new towns'.

The most beguiling images of a community governed by mathematical order are the surviving Renaissance representations of 'ideal cities'. None is more serenely beautiful than a painting from Urbino that is currently on show in Florence in the exhibition *L'uomo del Rinascimento (The Renaissance Man)*, which can be seen at the Palazzo Strozzi in Florence until 23 July. The exhibition centres on the architecture and writing of Leon Battista Alberti, author in 1435 of *On Painting*, the first book to expound the theory of linear perspective.

Many aspects of the ideal city from Urbino remain elusive. The artist and date are uncertain, and its original purpose is unclear. The most closely dated parallels are some doors, decorated with inlaid wooden panels of ideal cities, in the Ducal Palace at Urbino,

which were installed between 1474 and 1482. The ruler of Urbino, Federico da Montefeltro, was shaping Urbino itself into as much of an ideal city as the existing buildings would allow.

The painting in question was also part of a decorative ensemble, as its edges show that it was built into a piece of furniture, perhaps a chest, couch or bed. Infrared reflectography has revealed the huge effort that went into drafting the carefully conceived buildings in absolutely meticulous perspective, by disclosing part of the intricate carbon-based underdrawing. One surprise is that the central 'temple' was originally conceived as a structure with a flat façade and loggia below a triangular pediment. It might have been a governmental building, like those found at the centre of squares in Tuscan new towns from the late thirteenth century onwards. It was then transformed into a centralized building, almost certainly religious in function.

Everything speaks of an orderly society. The temple refers to the majesty of God's order, and the 'parish church' behind and to the right serves for daily devotions. The palaces, with their gracious arcades, shelter pedestrians and their owners in comfort and style. Rooftop loggias provide retreats

for well-ventilated relaxation, and the tiled piazza is clean and hygienic. And the beneficent rulers have provided two octagonal wells from which the citizens can draw clear water.

The whole is an essay in the geometry of perfect proportions and the good life. Only the people are missing: the ideal city is awaiting its ideal citizens. Will the city form them, shaping their values? Or have the people formed the city as their utopia?

Alberti himself, as the author of a treatise on the family, believed there was a natural order in the created world, and that we should adopt it in our lives. But he was realist enough, as was Federico, to know that human society needs the central imposition of rules.

Whether a city for contented dwellers can be planned scientifically, or whether it emerges organically from the free-market jungle, remains a mighty issue for town planners. When the Renaissance actually constructed planned cities, as they did at Urbino and Pienza, they somehow achieved a precarious balance between imposition and individuality. It is a difficult balance to contrive.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

PARTICLE PHYSICS

The first axion?

Steve Lamoreaux

For almost 30 years, the hunt has been on for a ghostly particle proposed to plug a gap in the standard model of particle physics. The detection of a tiny optical effect might be the first positive sighting.

Writing in *Physical Review Letters*¹, Emilio Zavattini and colleagues of the Italian PVLAS collaboration report that a magnetic field can be used to rotate the polarization of a light wave in a vacuum. Although this is the first experimental evidence for such an effect, there is a well-rehearsed, but controversial, explanation for it: the existence of a never-before-seen, chargeless, spinless and near-massless particle — the axion. Has the elusive axion finally allowed itself to be glimpsed?

The idea that static electric and magnetic fields alter the optical properties of matter, whether solid, liquid or gas, is not new. In 1845, Michael Faraday showed that the direction of polarization of light changes in a medium permeated by a magnetic field. Similar demonstrations followed: of the Cotton–Mouton and Kerr effects, for instance, in which the propagation speed of a light wave can be made to vary according to the orientation of its polarization with respect to an applied magnetic or electric field, respectively.

That static electromagnetic fields can be used to alter the properties of empty space might seem surprising. The culprit is quantum electrodynamics (QED), the quantum field theory that expresses the electromagnetic interaction in terms of the exchange of quanta of energy — photons — between particles. It was recognized in the early 1930s that the rules of QED allowed the creation and annihilation in an electromagnetic field of short-lived ‘virtual’ charged-particle pairs consisting of an electron and its antiparticle, the positron. The resulting ‘self-interaction’ of the field introduces a nonlinearity in the governing equations of electromagnetism (Maxwell’s equations) analogous to that induced by the presence of matter. The vacuum self-interaction is further complicated by the fact that, given the presence of sufficiently energetic photons, real (as opposed to virtual) electron–positron pairs can be created. Nevertheless, by 1936, vacuum analogues of various optical effects in matter had already been calculated^{2,3}.

So what other particles might the vacuum be hiding? In the late 1970s, a suitable candidate was postulated to solve the so-called strong CP problem. This problem pops up as a term in

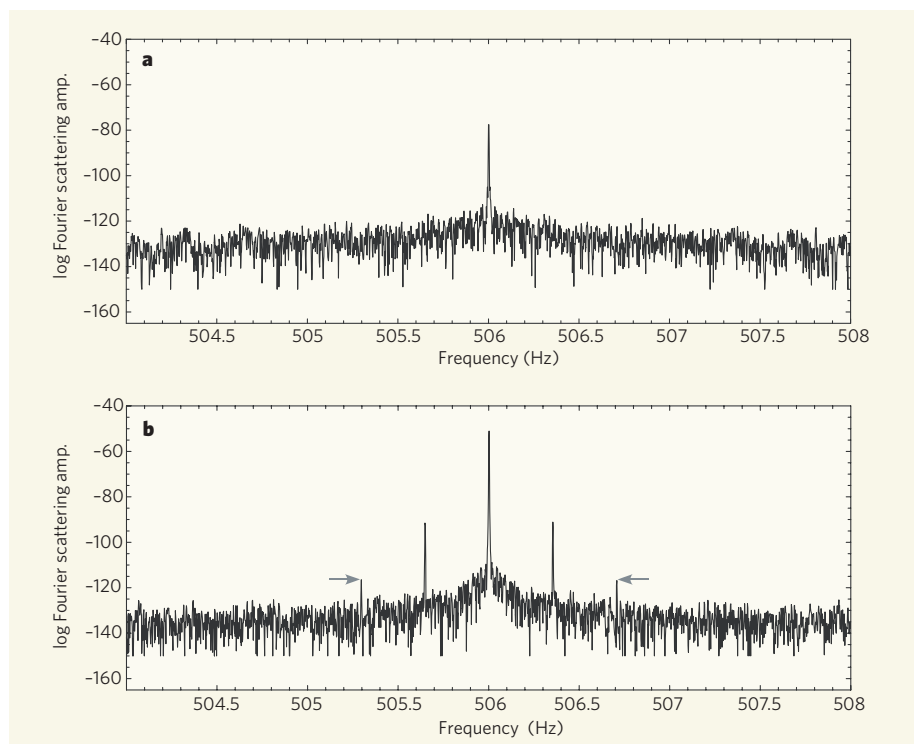


Figure 1 | Now you see it? Before (a) and after (b) Fourier amplitude spectra (logarithmic scale) of Zavattini and colleagues' vacuum polarization-rotation experiment¹. A sine wave generated at 506 hertz is modulated by the frequency of rotation, $\nu \sim 0.3$ Hz, of the experimental apparatus, which consists of a 1-metre-long interaction region traversing the bore of a 5.5-tesla magnet cooled cryostatically to 4.2 kelvin. **a**, With the magnetic field off, a central peak is observed, corresponding to the generated sine frequency. **b**, With the magnet on, side peaks are observed at frequencies displaced from the central peak by $\pm \nu$, corresponding to the magnet's rotation. Crucially, a signal displaced by $\pm 2\nu$ is also observed (arrows). This unexpected additional rotation of the light polarization vector could be the signal of axion production.

the field equations of quantum chromodynamics (QCD) — the equivalent of QED for the strong nuclear force, which binds together the innards of neutrons and protons. This term is not symmetric when, in a process involving the strong force, particles are replaced by antiparticles (charge conjugation; C) and the process is viewed in a mirror (parity inversion; P). The degree of this ‘symmetry breaking’ can be parametrized by an angle that, according to a limit obtained from measurements of the electric dipole moment of the neutron, is less than a billionth of a radian. Strong-force interactions covered by QCD

thus hardly seem to break CP symmetry at all.

This is an odd fact, because interactions involving the weak nuclear force (which can be incorporated with the electromagnetic force in QED) observably break CP symmetry. As there is no a priori reason for the discrepancy between these forces, Roberto Peccei and Helen Quinn proposed⁴ a mechanism implying the existence of another electrically neutral particle that would force the QCD symmetry-breaking angle — which could otherwise be as large as that in QED — to become zero. This particle cleaned up the strong CP problem, so Frank Wilczek dubbed it the ‘axion’, after a

laundry detergent that enjoyed limited commercial success in the United States⁵.

Although no axion has been detected in the 30-odd years since it was proposed, much can be said about its properties. Its mass (expressed in energy units based on the electronvolt, eV) must be greater than 1 μeV ; otherwise, it would have been too easily created in the Big Bang, so that there would be an amount of invisible 'dark matter' in the Universe that is incompatible with observations. On the other hand, its mass must be less than 0.01 eV; if it is not, an additional energy-loss mechanism would have been evident in the time course of the only supernova explosion, SN1987a, so far observed through modern telescopes⁶. Other astrophysical observations put a limit on the strength of the photon-axion interaction, dependent on the axion mass, of $0.6 \times 10^{-10} \text{ GeV}^{-1}$.

The axion is much lighter than an electron (at 511 keV, the electron weighs in at between 50 million and 500 billion times the axion mass), and can thus be produced more easily from a vacuum. Like the electron, it interacts with electromagnetic fields, so enhanced optical effects in a vacuum would also be expected. Such effects generally depend on the inverse fourth power of a particle's mass, so, even taking the upper limit of the axion mass, a factor of 10^{30} enhancement would be expected compared with the effect of an electron. This enhancement is, however, reduced by the much weaker coupling of the axion to the field.

It is just such an optical effect that the PVLAS collaboration claims to have seen¹ (Fig. 1). In their vacuum experiment, laser light of wavelength 1,064 nanometres — corresponding to a photon energy 1.2 eV greater than the maximum possible axion mass — propagated perpendicularly to an extremely strong (5 tesla) magnetic field 1 metre long. The light's polarization vector lay in the plane perpendicular to the direction of light propagation, and the authors varied its angle with respect to the magnetic field by turning the magnet at a constant rate around the laser beam.

Whenever the angle between the light field and the magnetic field was $\pm 45^\circ$ or $\pm 135^\circ$, the authors observed that the light polarization angle rotated with alternating sign and with a magnitude of 4×10^{-12} radian per metre of rotation path. This optical rotation angle, which becomes modulated at twice the magnet rotation frequency, corresponds exactly to what one would expect if the production of axions was attenuating the component of the light-field vector along the direction of the magnetic field. The inferred axion mass is about 1 meV — within the expected range — but the strength of the axion-photon interaction is, at about $5 \times 10^{-5} \text{ GeV}^{-1}$, far greater than expected. The net implication is that, if there is something there, it cannot be the QCD axion as circumscribed by astrophysics.

The PVLAS collaboration has done a credible investigation of possible spurious signals

that might explain its results. To reduce measurement errors and increase the experimental sensitivity, the authors magnified the rotation angle to about 0.1 microradian by repeatedly — around 44,000 times — passing the laser beam through the field region using a device known as an optical resonator. Also, the group calibrated the system and verified its efficient operation by introducing a low-pressure gas into the laser-beam vacuum region and confirming the Cotton-Mouton effect.

One worrisome point that has not been fully addressed, however, is the presence of a signal at the rotation frequency that is about 100 times larger than the putative axion signal. Although no correlation to the signal was found at twice the rotation frequency — the frequency at which the axion signal was identified — there could be a small component that is correlated, but that is masked by other random processes. That might be enough to generate a false signal. The presence of such a component might be due to inadvertent tilting of the apparatus, or the attraction of nearby ferromagnetic objects when the magnet is energized. Also still to be investigated is how the observed effect depends on the magnetic field strength, current reversal and laser wavelength.

The effect seen by the PVLAS collaboration is large enough for it also to be observed in other ways. In particular, a strong signal should be seen if laser light is polarized parallel to an applied static magnetic field and then blocked by an opaque barrier. Axions generated from the photons in the light would penetrate this

barrier easily because of their feeble interaction with matter, and a back-conversion of axions to photons could be accomplished by a second, identical magnetic field region behind it⁷.

Detecting such reconverted photons would be a much more robust signature of axions than determining the polarization state of light stored in an optical resonator. It had been thought that such an experiment would not be sufficiently sensitive because it requires a feeble conversion process to be applied twice. But as the large axion-photon coupling constant found by the PVLAS collaboration indicates a higher production rate of axions, one could imagine an axion 'factory' constructed of surplus powerful dipole magnets obtained from particle accelerators.

As befits the potentially revolutionary nature of the PVLAS result, the jury is still out. Such a direct verification would, however, propel it to a place among the most significant in the history of physics. ■

Steve Lamoore is at the Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. e-mail: lamore@lanl.gov

1. Zavattini, E. et al. *Phys. Rev. Lett.* **96**, 110406 (2006).
2. Heisenberg, W. & Euler, H. Z. *Phys.* **98**, 714–732 (1936).
3. Bialynicka-Birula, Z. & Bialynicki-Birula, I. *Phys. Rev. D* **10**, 2341–2345 (1970).
4. Peccei, R. & Quinn, H. *Phys. Rev. D* **16**, 1791–1797 (1977).
5. Wilczek, F. <http://nobelprize.org/physics/laureates/2004/wilczek-lecture.pdf> (2004).
6. Raffaeli, G. *Phys. Lett. B* **592** (Review of Particle Physics), 391–393 (2004).
7. Ansel'm, A. A. *Yad. Fiz.* **42**, 1480–1483 (1985).

CELL BIOLOGY

Divining cancer cell weaknesses

William G. Kaelin

Tumour cells tend to carry many gene mutations, but at a potential cost to their overall fitness. Studying the interactions between genes on a large scale could be a way of identifying the chinks in the tumour cell armour.

Cancer arises when the right combination of genes is mutated in a susceptible cell, much like lock tumblers falling into place. These mutations bestow various properties of malignancy upon the cell, such as independence from growth control and the ability to colonize other tissues. The fact that such cells contain multiple gene mutations tends to be viewed as problematic by developers of anticancer treatments — the typical thinking goes that each mutation is beneficial to the cancer cell and makes it harder. But the mutations probably come at a cost to the cell with respect to its ability to respond to certain situations. Unfortunately, we cannot yet predict how cancer-associated mutations might make a cell vulnerable, and there have been no unbiased methods for systematically discovering these

weaknesses. On page 106 of this issue, however, Louis Staudt and colleagues¹ describe an approach for identifying genes that become essential for the survival of cancer cells.

The idea that the deleterious consequences of a particular mutation might be revealed only under specific conditions is an old one. Not surprisingly, it has been most extensively explored in organisms such as yeast and fruit-flies, in which the genome can be manipulated easily and the consequent changes seen rapidly. Such investigations uncovered a gene-gene interaction, called 'synthetic lethality', that has potential significance for cancer drug discovery. Two genes (A and B, say) are said to be synthetically lethal if the cell containing them dies when both genes are mutated, but it can survive if either gene alone is mutated. In

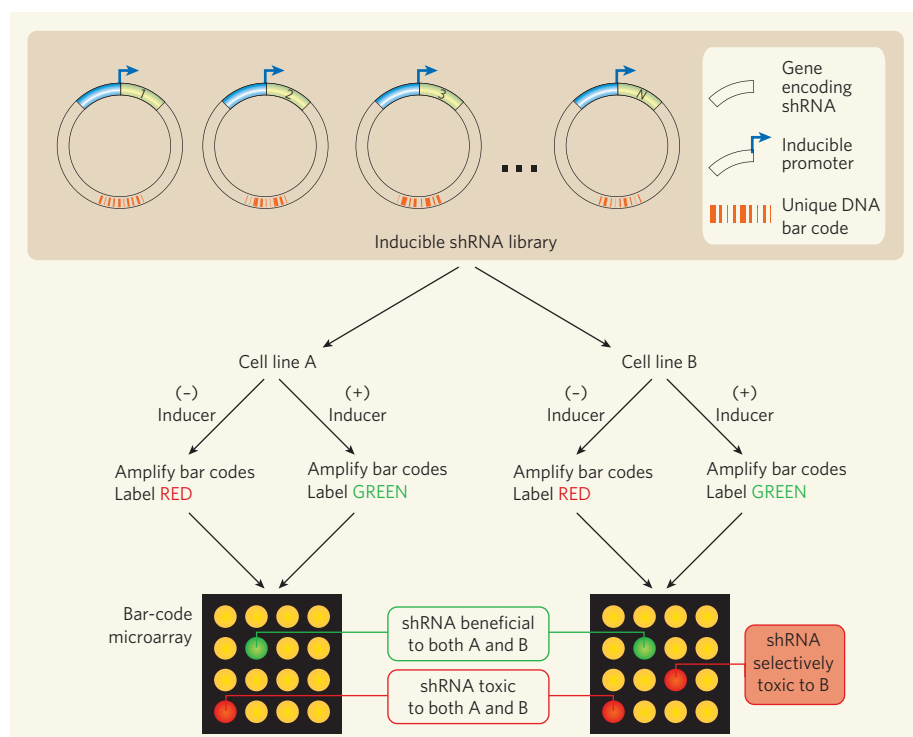


Figure 1 | Identifying shRNAs that impair cancer-cell fitness in a context-dependent manner. The scheme developed by Staudt and colleagues¹ uses a library of DNA vectors encoding shRNAs that block the expression of human genes. The expression of the shRNAs is controlled by a chemical inducer, and each vector contains a unique bar-code sequence that enables its subsequent abundance to be measured by a microarray. The library is introduced into two cell lines, 'A' and 'B', which could be two different cell lines, from different cancers say, or the same cell line grown under two different experimental conditions. Vector abundance gives an indication of cell fitness. On the microarray, abundances that increase or decrease when the shRNA gene is induced will appear green or red, respectively, and those that do not change will appear yellow (overlying red with green gives yellow).

perhaps the simplest case, *A* and *B* perform the same function and are uniquely redundant with respect to one another. Alternatively, *B* might be part of a pathway that can 'rescue' the pathway that is damaged by mutation of *A*. Studies in yeast, however, suggest that these possibilities are the tip of the iceberg — synthetic-lethal interactions seem to be unexpectedly common, although they are often hard to predict^{2,3}.

Hartwell *et al.*⁴ proposed that synthetic-lethal interactions could be used to develop safer and more effective anticancer drugs. In particular, they considered the case of tumour-suppressor genes, the protein products of which inhibit tumour growth, and which are often inactivated in cancers by mutations. A drug that inhibits the protein product of a gene that is synthetically lethal to a tumour-suppressor gene would, by definition, kill those cells in which the tumour-suppressor gene was inactivated, but not their normal counterparts.

This approach is conceptually appealing because it turns cancer-specific mutations into a liability for the cells that contain them. It also tackles two vexing problems in cancer drug discovery: how to kill cancer cells without harming normal cells, and how to tackle 'loss-of-function' mutations pharmacologically — most effective drugs cause a loss of function, rather than restoring a damaged function.

Synthetic lethality typically involves two loss-of-function mutations, but it also applies to gain-of-function mutations — those making an enzyme permanently active, say. For example, a gain of function in *A* might be synthetically lethal if it is not tolerated when *B* is inactive. In the jargon of cancer biology, genes that contribute to cancer when carrying gain-of-function mutations are termed proto-oncogenes, and one can envision other targets that might become vital to the cell specifically in the context of such proto-oncogene mutations.

But where does this leave the study of human cancer? So far, there are only a few reports of synthetic-lethal interactions involving human cancer genes, and we do not know enough about the genetic networks in cancer cells to predict such interactions. Although discovering synthetic-lethal interactions is fairly simple in yeast, most human cancer genes do not have relatives in yeast cells and the analogous experiments are much trickier, or nigh-on impossible, in human cells. So Staudt and colleagues developed a much-needed, high-throughput approach to discovering synthetic-lethal interactions in human cancer cells.

Their method is based on a technique called RNA interference, which involves the RNA molecules that act as a bridge between a gene and its product. In this technique, a 'short

interfering RNA' (siRNA) molecule is used to target a specific messenger RNA for destruction within the cell — halting production of its encoded protein. It is possible to perform large-scale genetic screens in human cells using libraries of genes that encode numerous siRNAs or 'short hairpin RNAs' (shRNAs, which are processed to siRNAs *in vivo*)⁵. Certain human shRNA libraries contain genes that are tagged with a unique DNA sequence, or 'bar code'^{6,7}. These bar codes can be used to identify the shRNAs that are selectively enriched under certain experimental conditions^{8,9}. However, identifying shRNAs that are depleted under certain conditions — which would be required for synthetic-lethal screens — has been problematic in human cells.

Staudt and colleagues¹ created a human shRNA library of DNA 'vectors' representing around 2,500 genes. Each vector contained a unique 60-base-pair bar code, and was designed such that shRNA expression could be switched on with a compound called doxycycline. The authors introduced the libraries into cells from two different types of diffuse large B-cell lymphoma — called activated B-cell-like (ABC) and germinal-centre B-cell-like (GBC) diffuse large B-cell lymphoma — on the assumption that these two types of cancer are driven by different mutations. The cells were then divided into two groups and either treated with doxycycline or left untreated. The abundance of the individual shRNA vectors was monitored using a DNA microarray consisting of probes that could detect the different bar codes in the library (Fig. 1). Vector abundance gives an indication of its effect on cell viability, or 'fitness', as abundance will increase as the cells divide and will decrease as cells die. So once a particular shRNA is induced by doxycycline, the effects of the reduction in the specific protein it targets can easily be monitored.

As expected, many shRNAs did not affect cellular fitness, or affected fitness negatively in both lymphomas. However, shRNAs targeting the genes *IKBKB*, *CARD11*, *MALT1* and *BCL10* were depleted after doxycycline addition only in the ABC cells. These shRNAs inhibited ABC proliferation selectively in further experiments, suggesting that *IKBKB*, *CARD11*, *MALT1* and *BCL10* are synthetically lethal to one or more mutations responsible for ABC. Alternatively, the normal cells that are the precursors of ABC might rely more heavily on these genes than do the progenitors of GBC because of heritable, non-genetic differences that persist after malignant conversion. These genes all encode agonists for the gene regulatory factor NF- κ B, so the results are consistent with an earlier finding that ABC is particularly sensitive to small-molecule NF- κ B antagonists¹⁰.

The approach described by Staudt and colleagues could be used to examine pairs of cell lines that differ only with respect to a mutation in a particular proto-oncogene or tumour-suppressor gene. In such a screen, a 'hit' would

be an shRNA vector that was depleted after shRNA induction in the mutant cells, but not their unmutated counterparts. Libraries of shRNAs should continue to improve, increasing genomic coverage and the degree of target inhibition, and so becoming ever more powerful tools in this endeavour. Sociologists have noted that some technological advances enable scientists to become 'communal harvesters' rather than 'hunter-gatherers'. With luck, the approach described by Staudt and colleagues will yield a bountiful harvest of cancer-drug targets that exploit weaknesses created by cancer-associated mutations. ■

William G. Kaelin is at the Howard Hughes Medical Institute, Dana-Farber Cancer Institute,

and Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.
e-mail: William_Kaelin@dfci.harvard.edu

1. Ngo, V. N. *et al. Nature* **441**, 106–110 (2006).
2. Sharom, J. R., Bellows, D. S. & Tyers, M. *Curr. Opin. Chem. Biol.* **8**, 81–90 (2004).
3. Kaelin, W. G. *Nature Rev. Cancer* **5**, 689–698 (2005).
4. Hartwell, L., Szankasi, P., Roberts, C., Murray, A. & Friend, S. *Science* **278**, 1064–1068 (1997).
5. Willingham, A. T., Deveraux, Q. L., Hampton, G. M. & Aza-Blanc, P. *Oncogene* **23**, 8392–8400 (2004).
6. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. *Nature Genet.* **14**, 450–456 (1996).
7. Hensel, M. *et al. Science* **269**, 400–403 (1995).
8. Berns, K. *et al. Nature* **428**, 431–437 (2004).
9. Paddison, P. J. *et al. Nature* **428**, 427–431 (2004).
10. Lam, L. T. *Clin. Cancer Res.* **11**, 28–40 (2005).

PLANETARY SCIENCE

A new spin on Saturn

David J. Stevenson

Measuring the rotation of a gaseous planet is no easy task. For Saturn, do observations of its magnetic field — which indicate that it is spinning more slowly than thought — mark a revolution in our understanding?

According to the International Astronomical Union (IAU), Saturn rotates once every 10 hours, 39 minutes and 22.4 seconds. Astronomical observations as far back as William Herschel's in the late eighteenth century¹ had suggested values for Saturn's rotational period of ten or so hours, but the IAU's seemingly precise value was defined by a periodicity in the kilometre-wavelength radio signal sent out by Saturn and detected by NASA's Voyager spacecraft in 1980. Surprising though it may seem, these radio emissions are still not well understood and — unlike radio emissions from pulsars or from Jupiter — turn out to be imprecisely periodic.

In this issue, Giampieri and colleagues (page 62)² report the observation, in data from the Cassini mission currently investigating the Saturn system, of a signal in Saturn's magnetic field of period about 10 hours 47 minutes. The authors suggest that this might be the planet's true rotation period.

So why should we care? One obvious reason is that rotation period is a fundamental property of a planet and tells us something about the conditions — such as the total angular momentum — when it formed. But that does not require a highly precise value. A less evident reason is that, with an equatorial radius some 10% greater than its polar radius, Saturn is the most distorted planet in the Solar System. Both the rotation and internal structure of a celestial object contribute to such an 'equatorial bulge', and the discrepancy of about 8 minutes (more than 1%) between the accepted and the new value for Saturn's rotational period will thus also affect our estimates of the size of its

likely inner core of rock and ice. Although the 'gas giant' Saturn is made mostly of hydrogen and helium, understanding the origin and evolution of such planets depends crucially on the nature and internal distribution of their minor constituents³. Saturn's rotation rate also provides the essential reference frame within which to talk about the dynamics of its atmosphere, and in particular the velocity of its very strong east–west winds. And this last point highlights a central issue: what does the rotation rate of a fluid planet actually mean?

The rotation rate of a planet such as Earth is conventionally defined to be that of its solid mantle, and can be measured to exquisite precision using geodetic methods such as the Global Positioning System. Tectonic movements of Earth's solid parts are, on average, slower than its rotation speed by 12 orders of magnitude, but are readily measurable. Larger effects arise from the interplay of tides and Earth's rotation, which causes an inexorable extraction of spin angular momentum, and the Moon to recede from Earth. Even larger, but fluctuating, effects arise from the exchange of angular momentum between Earth's solid regions and its fluid parts — the atmosphere, oceans and core. But the largest of these is still only a millionth of a hypothetical 1% saturnian change over the past 20 years.

Such a large variation in Saturn's rotation might be explained by its atmospheric features, in particular its strong easterly winds of up to 400 m s^{−1} relative to the IAU period. But a 1% difference in rotation rate of an atmospheric feature corresponds to a 'wind' of around 100 m s^{−1}, and, although Saturn's

winds may have changed, it is unlikely to have been by that much.

A more reliable measure of a fluid planet's spin than cloud patterns on its surface comes from its magnetic field. If this field is large and emanates principally from a dipole whose axis is tilted with respect to the axis of rotation, there will be a distinctive, readily detectable periodicity in the field and any resulting radio emissions. Because this magnetic field is generated deep down, its lines are embedded deeply in a large fraction of the planetary mass where large variable rotational motions should not occur. An east–west motion of just 1 m s^{−1} would, for instance, generate a toroidal magnetic field orders of magnitude larger than the observed dipole field, causing a Lorentz force to act and enforce uniform rotation again.

This argument agrees with our understanding of Earth. In Earth's liquid core, movements at depth are undetectable, but the movements at the top, in part revealed by magnetic-field changes over decades, indicate only tiny differential motions. Even the seismologically detected superrotation of Earth's inner core⁴ corresponds to only about one part in 10¹⁰ of Earth's total rotation.

Giampieri and colleagues² have found a stable periodicity in Saturn's magnetic field. Their observations are, however, incompatible with the expected inverse cubic dependence of field strength on distance from the planet that is expected from a tilted dipole. Saturn's magnetic field (Fig. 1) certainly lacks the substantial dipole tilt that we see so strikingly in Earth, Jupiter and, in a more complicated way, in Uranus and Neptune; but it would be astonishing if Saturn's field was exactly symmetric about the rotation axis. There is no reasonable model for field structure that can explain the observed behaviour solely through the internal field.

The observed periodicity might instead result from something that is a proxy for the internal rotation, for example a disturbance generated in Saturn's magnetosphere by a field anomaly (not necessarily dipolar) that is rotating with the core of the planet. The stability of the period would certainly be easiest to explain with the rotation of such a massive object. But although Saturn's deep interior would undoubtedly be a good clock, some region of smaller inertia would not necessarily be a bad one: the observed periodicity might be an indication of some wave-like or advected feature of the planet's ionosphere, and so be quite possibly unrelated to a rotation of the whole planet. In this case, it is curious that all the observed atmospheric motions rotate faster than the proposed rotation.

But the central puzzle would seem to be why Saturn behaves differently from Jupiter. Jupiter has a tilted dipole, a magnetic field that is much larger than Saturn's (even after taking account of the different-sized metallic region at its core), and winds that move both with and against the sense of the planet's rotation.

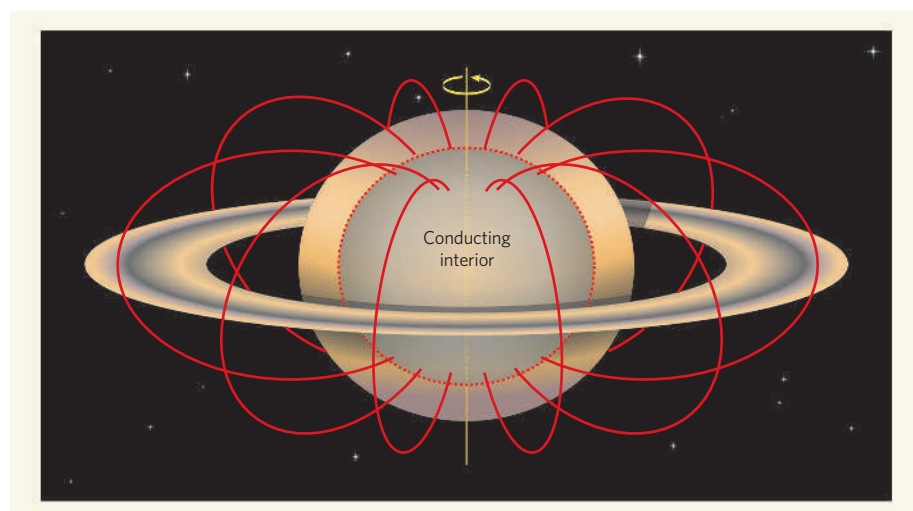


Figure 1 | Straight-up rotator. Saturn's magnetic field is almost exactly aligned with its rotation axis, and the dipole field lines emanate from a massive electrically conducting metallic hydrogen core (both axes are tilted by about 27° with respect to the plane of its orbit). Giampieri and colleagues' observation² of a periodic field variation, although not fully understood, could indicate that Saturn is rotating more slowly than had been assumed.

Owing to the much greater tilt of its axis of rotation (27° against 3°), Saturn has appreciable seasons, whereas Jupiter does not; perhaps we need a longer time base to appreciate fully the differences arising from this. (The orbital periods of the two planets are 29.5 years and 12 years, respectively.) However, this can only explain superficial differences. Maybe the Jupiter–Saturn differences are just another example of planetary diversity: the remarkable richness of planets in the Solar System alone

has arguably been the most striking outcome of the planetary exploration programme. ■ David J. Stevenson is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. e-mail: djs@gps.caltech.edu

1. Herschel, W. *Phil. Trans. R. Soc. Lond.* **84**, 48–66 (1794).
2. Giampieri, G., Dougherty, M. K., Smith, E. J. & Russell, C. T. *Nature* **441**, 62–64 (2006).
3. Guillot, T. *Science* **286**, 72–77 (1999).
4. Zhang, J. *et al. Science* **309**, 1357–1360 (2005).

MOLECULAR BIOLOGY

Chromosome guardians on duty

Paul Megee

Curiously, in cell division the proper separation of chromosomes into daughter cells needs set periods when they are stuck together. So how do they come apart at the right time and place? Their 'guardian spirits' intercede.

To avoid cell death or genetic diseases such as cancer, chromosomes must be transmitted to progeny cells with high fidelity. During cell division, chromosomes are duplicated to form sister chromatids, which must then be divided into two equal groups and separated into daughter cells (a process termed segregation). To prevent random segregation, sister chromatids are held together along their lengths by a ring-shaped protein complex called cohesin¹. At the cell-cycle stage known as anaphase, cohesins are cleaved and cohesion between sister chromatids dissolves, triggering their segregation. So the chromosomal association and dissociation of cohesin must be tightly regulated for proper segregation. In this issue, Kitajima *et al.* (page 46)² and Riedel *et al.* (page 53)³ describe how proteins known as

shugoshins — Japanese for 'guardian spirits' — and an associated regulatory enzyme temporally and spatially control the removal of cohesins from chromosomes.

The association of cohesins with the chromosome is particularly robust near the centromere⁴. This specialized region of the chromosome mediates assembly of the kinetochore complex that attaches the chromosome to the 'spindle', the cellular structure that pulls sister chromatids apart once cohesion has dissolved. Strong cohesion flanking the kinetochores is crucial, because it promotes the attachment of the two kinetochores on each chromatid pair to spindle fibres that originate from opposite sides of the dividing cell. It also opposes the splitting forces exerted by spindle fibres, preventing

premature separation of sister chromatids.

There are two types of cell division — meiosis, a specialized process that generates reproductive cells, and mitosis, which produces all other cells. Each has a pathway to deal with cohesin removal. In vertebrate mitosis, cohesins are detached from the chromosome arms first, in prophase (early mitosis), but the cohesins around the centromere are retained until the onset of anaphase (late mitosis) (Fig. 1, overleaf). At anaphase, an enzyme called separase cleaves the cohesin subunit known as Mcd1/Sccl, causing cohesins to dissociate from the centromeres⁵. How cohesins are removed from the chromosomal arms during prophase is less clear, other than it does not require separase and it is triggered by the phosphorylation (addition of a phosphate group) of a cohesin subunit called Sccl/SA (ref. 6).

In meiosis, DNA replication is followed by two successive rounds of chromosome segregation and cell division, producing four egg or sperm cells with half the chromosomal content of the other cells in the body (so that the fusion of two cells at fertilization gives the correct genetic complement). Maternally and paternally derived copies of chromosomes — referred to as homologues — are segregated in the first round, and sister chromatids are segregated in the second round. Before the first meiotic division, homologues usually undergo 'recombination', the reciprocal exchange of chromosomal segments. Recombination produces crossovers, which are physical links between homologues that can only be untangled by the release of sister-chromatid cohesion (Fig. 1). However, complete removal of cohesins at this stage would lead to random segregation of sister chromatids during the second division, so cohesin association persists in centromeric regions until the second division. At the first division, cohesins are removed in prophase by a pathway that is probably similar to that operating in mitosis⁷, and by the separase-mediated cleavage of Rec8, a meiosis-specific variant of Mcd1. Cohesin removal by these pathways seems to be enhanced by phosphorylation of cohesin subunits^{7,8}. As in mitosis, the removal of centromeric cohesins is dependent on separase.

How cohesin binding is stabilized specifically in centromeric regions is unclear. But hints came from the fruitfly protein MEI-S332, which, when inactivated, leads to defective centromeric cohesion⁹. MEI-S332 is the founding member of a family of proteins called shugoshins, whose localization to the kinetochore is essential for the persistence of centromeric cohesin^{10–12}. Shugoshin family members exist in organisms from yeast to humans, implying that there is an evolutionarily conserved mechanism to protect centromeric cohesin.

To determine how shugoshins protect centromeric cohesin, Kitajima *et al.*², Riedel *et al.*³ and Tang *et al.*¹³ set out to identify proteins that interact with shugoshin in mitotic and

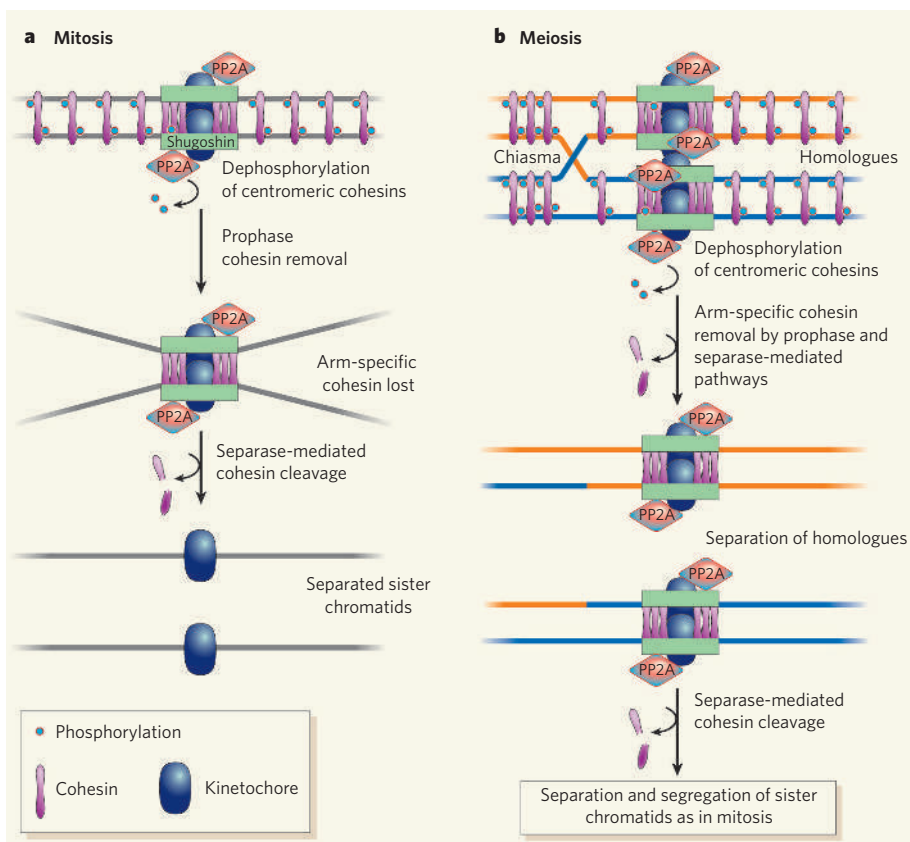


Figure 1 | Cohesins in cell division. **a**, During mitosis, phosphorylation of a cohesin subunit promotes removal of cohesins by a pathway in the prophase stage of cell division. Shugoshin mediates recruitment of PP2A to kinetochores, the protein complexes at the chromosome centromeres. The PP2A enzyme dephosphorylates the cohesin subunits and so maintains cohesion at the centromere. At the onset of the anaphase stage of division, cleavage of the Mcd1 cohesin subunit removes cohesins from centromeric regions, allowing sister chromatids to separate. **b**, In meiosis, maternally and paternally derived homologues undergo recombination during the first division, producing crossovers (chiasmata). Cohesin-subunit phosphorylation promotes cohesin removal by two mechanisms: a prophase pathway that is probably similar to that in mitosis, and the separase-mediated cleavage of meiotic cohesin subunit Rec8. The localization of PP2A to centromeric regions by shugoshins results in the dephosphorylation of centromeric cohesin subunits, making them resistant to removal by the prophase and separase-mediated pathways. At the second meiotic division, centromeric cohesins are removed by separase-mediated cleavage.

meiotic cells. Each group identified an enzyme called protein phosphatase 2A (PP2A) that dephosphorylates certain protein residues. PP2A is composed of catalytic, scaffolding and regulatory (B) subunits, and although several subfamilies of PP2A regulatory subunits exist, the shugoshins interacted only with the subset of PP2A that contained B' regulatory subunits^{2,3}. Moreover, despite the fact that PP2A is found in many places in the cell, Kitajima *et al.*² and Riedel *et al.*³ show that the B'-containing PP2A forms are found preferentially in centromeric regions, where they co-localize with shugoshins.

Given that phosphorylation of cohesin subunits triggers cohesin removal⁶, all three research groups^{2,3,13} propose that shugoshins protect centromeric cohesion in both mitotic and meiotic cells by recruiting PP2A to centromeres, where it stabilizes centromeric cohesin by dephosphorylating cohesin subunits. Consistent with this model was the dramatic increase in premature chromatid separation and chromosome mis-segregation

that was observed when either PP2A activity was eliminated in yeast, or PP2A protein levels were reduced in human cells^{2,3,13}. Furthermore, Kitajima *et al.*² showed that in cells during late mitosis, chromosome-bound cohesins (which are mostly centromeric) were phosphorylated to a lesser extent than were free cohesins (presumably those formerly associated with chromosome arms). So it seems likely that PP2A protects centromeric cohesin by reducing cohesin-subunit phosphorylation in centromeric regions.

But is getting PP2A to the centromeres sufficient to prevent cohesin removal? Kitajima *et al.*² and Riedel *et al.*³ artificially tethered PP2A to chromosomes independently of shugoshins. Under these conditions, cohesins were retained on the chromosomes and the cells either failed to undergo nuclear divisions or aberrantly segregated their chromosomes. This supports the idea that the main function of shugoshins is to recruit PP2A to the centromere^{2,3}. However, the depletion of one shugoshin protein in cells expressing a second

family member induced premature chromatid separation even though PP2A was tethered to the centromere, suggesting that shugoshins may have additional roles in protecting centromeric cohesion².

In a related study published online in *Nature*, Brar *et al.*¹⁴ investigated whether the phosphorylation of the meiotic cohesin subunit Rec8 is involved in the stepwise loss of cohesins during meiosis. Previous results showed that depletion of the enzyme thought to phosphorylate Rec8 leads to defects in Rec8 cleavage and entry into anaphase⁸. And consistent with this, Brar *et al.* found that cells with a mutant version of Rec8 in which phosphorylation was reduced had difficulties initiating the first meiotic division because of a defect in the separase-mediated cleavage of Rec8. The requirement for Rec8 phosphorylation was abolished, however, if there was no recombination between homologues. Without recombination-mediated crossovers, the removal of cohesins from chromosome arms was not required to separate homologues, and cohesins persisted on chromosomes until the second meiotic division.

So the two meiotic divisions have different requirements for Rec8 phosphorylation in triggering the separase-mediated cleavage of Rec8 — Rec8 phosphorylation is important for cleavage in the first meiotic division, but less so in the second division. Surprisingly, it seems that recombination itself might contribute to the ordered removal of cohesins from chromosome arms. In addition, separase-mediated cleavage of the mutant Rec8 did occur when shugoshin was depleted, leading Brar *et al.*¹⁴ to conclude, as have Kitajima *et al.*², that shugoshins may have other duties in addition to protecting centromeric Rec8 from phosphorylation.

These four studies have made significant progress in elucidating the role of shugoshins in protecting centromeric cohesin, but they also raise questions. For example, several organisms, including humans, have multiple shugoshin family members that have different expression patterns during mitotic and meiotic cell divisions, raising the likelihood that specific family members have different roles in protecting centromeric cohesin. Determining precisely how shugoshins function as a chromosome segregation nexus, coordinating both centromeric cohesion and other potential roles, will be fundamental to our understanding of the molecular mechanisms of cell division. ■ Paul Megee is in the Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, Mail Stop 8101, PO Box 6511, Aurora, Colorado 80045, USA. e-mail: paul.megee@uchsc.edu

- Haering, C. H., Lowe, J., Hochwagen, A. & Nasmyth, K. *Mol. Cell* **9**, 773–788 (2002).
- Kitajima, T. S. *et al.* *Nature* **441**, 46–52 (2006).
- Riedel, C. G. *et al.* *Nature* **441**, 53–61 (2006).
- Weber, S. A. *et al.* *PLoS Biol.* **2**, 1340–1353 (2004).
- Weizenegger, I. C. *et al.* *Cell* **103**, 399–410 (2000).
- Uhlmann, F., Lottspeich, F. & Nasmyth, K. *Nature* **400**, 37–42 (1999).

7. Yu, H.-G. & Koshland, D. *Cell* **123**, 397–407 (2005).
8. Lee, B. H. & Amon, A. *Science* **300**, 482–486 (2003).
9. Kerrebrock, A. W., Miyazaki, W. Y., Birnby, D. & Orr-Weaver, T. L. *Genetics* **130**, 827–841 (1992).
10. Kitajima, T. S., Kawashima, S. A. & Watanabe, Y. *Nature* **427**, 510–517 (2004).
11. Marston, A. L., Tham, W.-H., Shah, H. & Amon, A. *Science* **303**, 1367–1370 (2004).
12. Rabitsch, K. P. *et al. Curr. Biol.* **14**, 287–301 (2004).
13. Tang, Z. *et al. Dev. Cell* doi:10.1016/j.devcel.2006.03.010 (2006).
14. Brar, G. A. *et al. Nature* doi:10.1038/nature04794 (2006).

MATERIALS SCIENCE

Polymers show they're metal

Richard Friend

Although certain polymers have long been known to conduct electricity, they seemed to differ from metals in other electronic and optical properties. A new form of polymer turns that relation on its head.

The fact that the polymer polyacetylene reaches metallic levels of conductivity when chemically doped caused quite a stir when it was first reported¹ in 1977. This was not just because conducting polymers might be of great practical use, but also because the underlying physics came as a surprise: the apparently free movement of electrons along the carbon-based backbone of a polymer chain was not widely anticipated. The discovery attracted the attention of many chemists, materials scientists and physicists, and by the mid-1990s several practical conducting polymeric materials had emerged.

Despite their metal-like conductivities, these polymers did not pass muster as metals in other important respects. On page 65 of this issue, Lee *et al.*² supply missing evidence that supports the prevailing view that it is structural disorder, which is very hard to eliminate in polymeric materials, that prevents making real metals of polymers^{3,4}. The authors' doped, conducting form of polyaniline, synthesized to produce a more ordered material than was previously available, shows both the electrical and the optical properties required of a true metal.

Although the electrical conductivity of this form of polyaniline is, at about 10^3 siemens per centimetre at room temperature, not particularly high (the conductivity of copper at the same temperature is, for instance, around 6×10^5 S cm⁻¹), it increases steadily as temperature is lowered, and by 4 kelvin has risen by a factor of about 2.5. Such behaviour is a hallmark of a metal: the conduction electrons that carry the electrical current within a metal travel as free particles, and it is the rate at which these electrons are scattered that limits conductivity. Because the dominant source of scattering is the thermal vibrational motions of the atoms, as temperature falls, scattering is frozen out, and so conductivity increases. At very low temperatures, the conductivity saturates at a level set by scattering due to residual disorder.

Another defining feature of a metal is its optical reflectivity, which gives, for instance, silver and aluminium their familiar shine. Light is reflected from a metallic surface when

its frequency lies below the characteristic oscillation frequency (the 'plasma frequency', ω_p) of electrons in a metal. (This frequency is given by the expression $\omega_p^2 = ne^2/\epsilon m$, where n is the number density of conduction electrons, m the electron mass and ϵ is the background dielectric constant.) In this case, conduction electrons are able to move fast enough to screen out incident electromagnetic radiation and, mathematically, the real part of the complex dielectric response function of a metal is negative.

For metals with high densities of conduction electrons such as silver and aluminium, the plasma frequency lies above the frequency band of visible light, in the ultraviolet region of the electromagnetic spectrum. For Lee and colleagues' conducting form of polyaniline², however, the density of conduction electrons is much lower, and the plasma frequency is deep in the infrared, below visible-light frequencies. Polyaniline does not therefore seem to reflect visible light, as a metal does; nevertheless, its optical reflectivity in the infrared very accurately fits the model for a simple metal, and, critically, the real part of its dielectric response function remains negative down to the lowest frequencies.

In fact, the surprise is just how like an ordinary metal polyaniline turns out to be. Modelling its optical reflectivity assuming that it is a metal produces a realistic estimate of the number of conduction electrons (about one for every two aniline groups in the polymer), requires no adjustment to take into account the electron mass, and gives a value for d.c. conductivity that matches the directly measured value.

This finding pitches the metallic model against a different model, loosely referred to as the 'polaron model'⁵, that over the past two decades has been the favoured explanation of polymer conductivity. This model starts from the fact that, without chemical doping, polymers such as polyacetylene and polyaniline are semiconductors (as such, they are extensively used as the active component in transistors, light-emitting diodes and photovoltaic diodes⁶). When a small amount of dopant

electronic charge is added to the undoped polymer, the electronic charges on the polymer chain are localized. Thermal excitation is required to move charge from one site to the next, and so conductivity drops as temperature falls. Ironically, the localization that hinders conduction is substantially the handiwork of the agents that produced the charge: the chemical dopants are partially ordered in sites adjacent to the polymer chains (so-called 'charge transfer complexes') and their full-blown ionic charges provide random potentials that confine the electrons.

Localization is enhanced by the rearrangement near charges of covalent bonds on the polymer chain. This process produces a lower energy state for electrons to occupy, and the resultant coupling of the local structure to the electronic charge can be regarded as a particle-like entity, known as a polaron. It also causes a set of characteristic optical signatures, including vibrational modes of the polymer structure at frequencies between 15 and 60 terahertz and electronic transitions to states within the semiconductor gap, typically at frequencies in the near infrared.

Although there is plenty of spectroscopic evidence to support the polaron model for partially doped materials, polaronic features seem insignificant in Lee and colleagues' more ordered doped polyaniline². The transition from polaronic conductor to simple metal — from a material whose conductivity, at low temperatures, decreases with falling temperature to one whose conductivity increases — is surprising. The explanation must lie in the way that conduction electrons are able to screen out the local potentials that stem from both polaron geometrical relaxation and structural disorder. This process is self-perpetuating: as the potentials become screened, the conduction electrons are themselves able to become more delocalized, and thus better able to screen.

This phenomenon has parallels in some classes of inorganic materials such as 'colossal magnetoresistance' manganites that can show huge changes in conductivity as their magnetic order changes, and in which the role of polaronic interactions is similarly important⁷. The implications for the field of conducting polymers are bitter-sweet: although the demonstration of a polymer acting as a true metal is a real milestone, it sets a limit on the level of conductivity we can expect from such materials. ■

Richard Friend is at the Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK. e-mail: rhf10@cam.ac.uk

1. Chiang, C. K. *et al. Phys. Rev. Lett.* **39**, 1098–1101 (1977).
2. Lee, K. *et al. Nature* **441**, 65–68 (2006).
3. Kohlman, R. S. & Epstein, A. J. in *Handbook of Conducting Polymers* (eds Skotheim, T. A., Elsenbaumer, R. L. & Reynolds, J. R.) 85–121 (Dekker, New York, 1998).
4. Menon, R., Yoo, C. O., Moses, D. & Heeger, A. J. in *Handbook of Conducting Polymers* (eds Skotheim, T. A., Elsenbaumer, R. L. & Reynolds, J. R.) 27–84 (Dekker, New York, 1998).
5. Heeger, A. J. *et al. Rev. Mod. Phys.* **60**, 781–850 (1988).
6. Malliaras, G. & Friend, R. *Phys. Today* **58**, 53–58 (2005).
7. Littlewood, P. & Kos, S. *Nature* **438**, 435 (2005).

BRIEF COMMUNICATIONS

Complex call production in the túngara frog

An intricate vocal anatomy underlies the subtleties of this animal's melodic mating calls.

Animals' sound-producing organs often act as an integrated whole — particular vocal structure are not directly associated with the creation of discrete syllables^{1–4}. But here we show that the 'chuck' of the 'whine-chuck' mating call of the túngara frog, *Physalaemus pustulosus*, is caused by a fibrous mass attached to the vocal folds; the chuck is eliminated by removal of this structure, although the frog still tries to produce the sound. Sexual selection affects the acoustic complexity of the frog's call^{5–7}, so evolution may have shaped this unusual vocalization, which is akin to the two-voiced song of songbirds^{6,8}.

Calls of túngara frogs consist of a whine followed by up to seven chucks. The whine is a long-frequency sweep, whereas the chuck is a shorter sound whose fundamental frequency is half that of the whine (for details, see supplementary information). Females prefer a whine-chuck call to a whine alone (probability of 0.856; 3,135 versus 527; exact binomial, $P < 0.000001$).

In *P. pustulosus* and some close relatives, the vocal folds support a fibrous mass that extends into the bronchus (see supplementary information). The size of this mass correlates with the presence or absence of chuck-like call components among species within the genus¹, within the *P. pustulosus*

species group⁹ and between populations of the túngara frog's sister species, *P. petersi*¹⁰.

To investigate whether the fibrous mass could be responsible for production of the chuck sound, we recorded calls of male túngara frogs in the field ('pre-treatment' condition). We then anaesthetized the frogs, surgically removed the posterior elongation of the fibrous mass, and recorded their calls after they had recovered ('post-treatment' condition). A sham control replicated all aspects of the surgery apart from ablation of the mass. An additional control omitted surgical manipulation but replicated the feeding, housing and handling used in the other treatments.

As the fundamental frequency of the chuck is half that of the whine, its odd harmonics do not coincide with the whine's harmonics, although its even harmonics do (see supplementary information). We predicted that ablation of the fibrous mass would reduce the amount of energy in odd relative to even harmonics of the chuck, but would not influence the spectral features of the whine or the increased amplitude at the call's end.

The relative energy in odd versus even harmonics of the chuck was significantly reduced after surgical treatment ($F_{2,17} = 21.77$, $P < 0.001$; Fig. 1a, b). The harmonic-energy ratio was significantly lower than in either of

the two controls (Tukey post-hoc test; all comparisons, $P < 0.001$); the two controls did not differ from one another (Tukey test, $P = 0.538$). There was no treatment effect in the same analysis of the whine ($F_{2,17} = 0.891$, $P = 0.429$).

Although the chuck is diagnosed by its fundamental frequency relative to that of the whine, it also increases the amplitude of the end of the call. An increase in amplitude is therefore evidence that the male is attempting to produce a chuck. As predicted, there was no treatment effect on the relative amplitude at the end of the call ($F_{2,21} = 0.287$, $P = 0.754$; Fig. 1c, d). Males without the posterior elongation of the fibrous mass therefore still increase call amplitude at the end of the whine, as they would when producing a chuck, but the spectral structure of the chuck is missing. Simply put, males without an elongated fibrous mass on their vocal folds try to produce chuck sounds but are unable to do so.

M. Gridi-Papp*[§], A. S. Rand†, M. J. Ryan‡

*Departamento de Zoologia, Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, São Paulo 13506-900, Brazil

†Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Panama

‡Section of Integrative Biology, University of Texas, Austin, Texas 78712, USA

e-mail: mryan@mail.utexas.edu

§Present address: Department of Physiological Science, University of California, Los Angeles, California 90095, USA

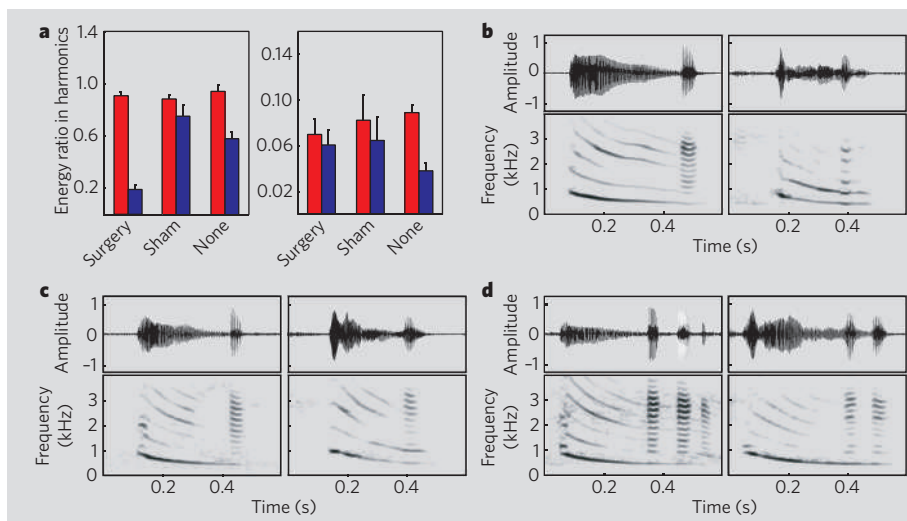


Figure 1 | Effect of different treatments on call harmonics of the túngara frog, *Physalaemus pustulosus*.

a, Relative amplitude of odd (red bars) to even (blue bars) harmonics in each surgical-treatment group for 'chuck' (left panel) and 'whine' (right panel) call components; means and standard errors are shown. Surgery, $n = 8$; sham, $n = 6$; none, $n = 6$. **b–d**, Representative calls (top, relative amplitude of waveforms; bottom, spectrogram) by an individual male in pre- (left panels) and post-treatment (right panels) conditions after **b**, ablation of the 'chuck'-associated fibrous mass; **c**, sham surgery; and **d**, no surgery.

1. Drewry, G., Heyer, W. R. & Rand, A. S. *Copeia* **1982**, 636–645 (1982).
2. Suthers, R. A. *Nature* **347**, 473–477 (1990).
3. Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. *Nature* **395**, 67–71 (1998).
4. Fitch, W. T. *Anim. Behav.* **63**, 407–428 (2002).
5. Ryan, M. J. *Science* **209**, 523–525 (1980).
6. Ryan, M. J. *The Túngara Frog: A Study in Sexual Selection and Communication* (Univ. Chicago Press, Chicago, 1985).
7. Ryan, M. J., Fox, J. H., Wilczynski, W. & Rand, A. S. *Nature* **343**, 66–67 (1990).
8. Greenwalt, C. H. *Bird Song: Acoustics and Physiology* (Smithsonian Institution Press, Washington DC, 1968).
9. Ryan, M. J. & Drewes, R. C. *Biol. J. Linn. Soc.* **40**, 37–52 (1990).
10. Boul, K. E. & Ryan, M. J. *Copeia* **2004**, 624–631 (2004).

Supplementary information accompanies this communication on Nature's website.

Received 5 January; accepted 27 March 2006.

Competing financial interests: declared none.
doi:10.1038/441038a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

THEORETICAL BIOLOGY

Comparing models of species abundance

Arising from: I. Volkov, J. R. Banavar, F. He, S. P. Hubbell & A. Maritan *Nature* **438**, 658–661 (2005)

Ecologists are struggling to explain how so many tropical tree species can coexist in tropical forests, and several empirical studies have demonstrated that negative density dependence is an important mechanism of tree-species coexistence^{1,2}. Volkov *et al.*³ compare a model incorporating negative density dependence with a dispersal-limited neutral model⁴ and claim that each predicts six empirical species-abundance distributions of tropical-tree communities equally well. However, we show here that their main conclusion is premature: when the two models are compared in an improved analysis, we find that the dispersal-limited model outcompetes the density-dependent model in all six cases. Hence, although density dependence is certainly an important diversity-maintaining mechanism, our improved approach indicates that the dispersal-limited model provides a more parsimonious explanation of empirical species-abundance distributions.

Volkov *et al.* compare model performances by applying a maximum-likelihood method to empirical species-abundance distributions. They use an approximate-likelihood function⁵, but an exact likelihood is now available for the dispersal-limited model that is based on a proper sampling theory^{6,7}. We applied the new maximum-likelihood method to the data sets used by Volkov *et al.* and based our model comparison on the difference in the models' AIC, defined as twice the number of parameters in the model minus twice its maximum log-likelihood⁸.

We found that the dispersal-limited model performed significantly better than the density-dependent model for all six forests (Table 1). The probability that the density-dependent model performed better than the dispersal-limited model (mathematically defined as the Akaike weight⁸) was always less than 1%. Even a dispersal-unlimited neutral model performed better than the density-dependent model in two cases out of six. Thus the claim by Volkov *et al.* that density dependence gives an equally sufficient mechanistic explanation for species-abundance distributions, in addition to and independent of dispersal limitation, is unsubstantiated.

How can one test the claim of Volkov *et al.* that alternative models cannot be distinguished, in principle or in practice, given just the species-abundance distributions? Volkov *et al.* support their assertion by arguing that the ratio $\hat{r}_n$, computed from the expected number of species with a given abundance³, looks the same for all models when plotted against n . However, they do not convincingly show that these graphs are

Table 1 | Comparison of species-abundance models

Site	S	J	Θ	m	L_1	L_2	L_3	Model ranking	w_2	w_3
BCI	225	21,457	47.7	0.093	−308.7	−315.0	−318.8	1 > 2 > 3	0.0018	0.0001
Yasuni	821	17,546	204.2	0.429	−297.2	−303.6	−307.6	1 > 2 > 3	0.0017	0.0001
Pasoh	678	26,554	190.9	0.093	−359.4	−365.3	−392.5	1 > 2 > 3	0.0027	0.0000
Korup	308	24,591	52.7	0.547	−317.0	−323.1	−318.7	1 = 3 > 2	0.0022	0.3318
Lambir	1,004	33,175	285.6	0.115	−386.4	−391.2	−437.9	1 > 2 > 3	0.0090	0.0000
Sinharaja	167	16,936	436.8	0.0019	−252.9	−258.5	−253.8	1 = 3 > 2	0.0037	0.2928

The performance of three species-abundance models are compared in six large tropical forest dynamic plots containing J sampled individuals and S species. Parameters Θ and m are the maximum-likelihood estimates in model 1. L_1 , L_2 and L_3 represent the maximum log-likelihood values for the respective models. Model ranking was based on the Akaike weights, w_i , the probability that model i is more likely than the model with the lowest AIC⁸.

Methods. Model 1 is the neutral model with dispersal limitation of ref. 4, model 2 is the density-dependence model of ref. 3, model 3 is like model 1 but with no dispersal limitation (ref. 4). In models 1 and 3, likelihoods are calculated based on ref. 6. In model 2, likelihoods are calculated based on ref. 5 (see ref. 3). The AIC is defined as $2(p - L_i)$ where p is the number of parameters ($p = 2$ in models 1 and 2, and $p = 1$ in model 3). The Akaike weights were computed by $w_i = \exp(-\Delta_i/2) / [1 + \exp(-\Delta_i/2)]$, where Δ_i is the difference of AIC between model i and the model with the lowest AIC.

statistically indistinguishable. Even if they had, this would still be a weak test of model equivalence. Proving that two species-abundance models are mathematically equivalent requires construction of the multivariate distribution for species abundances under the density-dependent model, as well as evidence that this distribution is mathematically equal to that derived in ref. 6. If so, the two models would indeed be indistinguishable. If the models are not mathematically equivalent, they may still be statistically equivalent in most practical situations. Proving statistical equivalence requires a case-by-case comparison of the models' exact likelihood functions, as outlined here.

Our analysis provides evidence in favour of the dispersal-limited model, and contradicts the claim of Volkov *et al.*³. However, the authors' density-dependent model in its current formulation lacks a suitable sampling theory with an exact-likelihood function, and hence precludes any definitive answer about the fundamental or practical equivalence of alternative models. Although Volkov *et al.* have made a useful contribution by emphasizing the value of simple models in ecology, further theoretical developments aimed at exploring

the contribution of density dependence to patterns of species diversity are needed^{9,10}.

Jérôme Chave*, David Alonso†, Rampal S. Etienne‡

*Laboratoire Evolution et Diversité Biologique, UMR5174 CNRS, Université Paul Sabatier, 31062 Toulouse, France
e-mail: chave@cict.fr

†Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, USA

‡Community and Conservation Ecology Group, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

1. Harms, K. E., Wright, S. L., Calderón, O., Hernandez, A. & Herre, E. A. *Nature* **404**, 493–495 (2000).
2. Packer, A. & Clay, K. *Nature* **404**, 278–281 (2000).
3. Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. *Nature* **438**, 658–661 (2005).
4. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography*. (Princeton Univ. Press, Princeton, New Jersey, 2001).
5. Alonso, D. & McKane, A. J. *Ecol. Lett.* **7**, 901–910 (2004).
6. Etienne, R. S. *Ecol. Lett.* **8**, 253–260 (2005).
7. Etienne, R. S. & Alonso, D. *Ecol. Lett.* **8**, 1147–1156 (2005).
8. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference* 2nd edn (Springer, New York, 2002).
9. Engen, S. & Lande, R. *Math. Biosci.* **132**, 169–183 (1996).
10. Diserud, O. & Engen, S. *Am. Nat.* **155**, 497–511 (2000).

doi:10.1038/nature04826

THEORETICAL BIOLOGY

Volkov et al. reply

Replying to: J. Chave, D. Alonso and R. S. Etienne *Nature* **441**, doi:10.1038/nature04826 (2005)

We have demonstrated that information on relative species abundance (RSA) cannot, without additional information, be used to discriminate among biological explanations for different RSA patterns¹; but Chave *et al.* claim that our conclusion is premature². Here

we show that their analysis was not carried out in a consistent manner and that density dependence gives an equally valid mechanistic explanation for RSA patterns in addition to, and independently of, dispersal limitation.

Consider a very simple model of immigra-

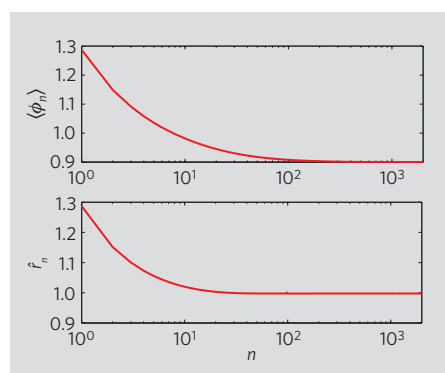


Figure 1 | Species abundance. Top, plot of $\langle \phi_n \rangle$, the average number of species with abundance n , against n for the model described in ref. 2, using the biodiversity parameter $\Theta = 50$, the immigration parameter $m = 0.1$ and a population of 20,000. Bottom, plot of $\hat{r}_n$, the density-dependent birth rate–death rate ratio, against n , deduced from the RSA data.

tion, in which the birth and death rates for a species with n individuals is given by $b_n = b^*n + c^*$ and $d_n = d^*n$, respectively, or the equivalent per capita rates are $b_n/n = b^* + c^*/n$ and $d_n/n = d^*$, where b^* and d^* represent the intrinsic per capita birth and death rates, respectively, and c^* is a constant that captures the effects of immigration from a surrounding metacommunity that has a uniform distribution of species. This model can equally well be envisaged as arising not from immigration but from a density-dependent birth mechanism, so it is impossible to determine the biology

underlying the resulting RSA data without additional information.

As a second example, consider the dispersal limitation model studied by Volkov *et al.*³ using the biodiversity parameter $\Theta = 50$ and the immigration parameter $m = 0.1$. Figure 1 shows the RSA data for this model obtained through the analytical expression derived in ref. 3, as well as a graph of the density-dependent birth–death rate ratio $\hat{r}_n$ plotted against n , deduced¹ from this RSA data ($\hat{r}_n = [(n+1)/n] [b_n/d_{n+1}] = [(n+1)/n] [\langle \phi_{n+1} \rangle / \langle \phi_n \rangle]$, where $\langle \phi_n \rangle$ represents the mean number of species with abundance n).

Assume now that we were given the RSA data (Fig. 1, top) and asked what we could learn from it. Clearly, the dispersal-limitation model from which the data were derived provides a perfect description of the data, so it might be tempting to conclude that the mechanism underlying the data is in fact dispersal limitation with the selected model parameters. However, a symmetric density-dependence mechanism that gives $\hat{r}_n$ versus n (Fig. 1, bottom) would also yield exactly the same RSA data. Furthermore, $\hat{r}_n$ does not constrain the individual birth and death rates, but only their ratio. This provides even greater flexibility in the range of models and mechanisms that are able to fit the data exactly.

Turning to the narrower issue of the comparison between two simplified models, the table presented by Chave *et al.*² is misleading because their analysis is inconsistent. Our assumption¹ was that the RSA data were a

measure of $\langle \phi_n \rangle$, the average RSA, whereas Chave *et al.*² consider the multivariate probability distribution for ϕ_n and maximize it for the snapshot in question. To compare the two models consistently, we computed $\langle \phi_n \rangle$ by generating 100,000 realizations of each plot with the estimates by Chave *et al.*² of the parameters Θ and m . The results for four forests (forests with a large value of m are very slow in their convergence) shown in Table 1 do not support the claim by Chave *et al.*².

In summary, any RSA data set is exactly equivalent to an effective symmetric density-dependence model. The key unknown is what led to this effective density dependence: did it arise from a dispersal-limitation model, by some other mechanism, or is there intrinsic symmetric density dependence at play? It is not possible, in principle or in practice, to distinguish between these options given just the RSA data¹.

Igor Volkov[†], Jayanth R. Banavar^{*},
Fangliang He[‡], Stephen P. Hubbell[§]||,
Amos Maritan[¶]

^{*}Department of Physics, 104 Davey Laboratory, and [†]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA
e-mail: jayanth@phys.psu.edu

[‡]Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1, Canada

[§]Department of Plant Biology, The University of Georgia, Athens, Georgia 30602, USA

^{||}The Smithsonian Tropical Research Institute, Box 2072, Balboa, Panama

[¶]Dipartimento di Fisica 'G. Galilei', Università di Padova, 35131 Padova, Italy

Table 1 | Model comparisons

Plot	L_1	L_2	L_3
BCI, Panama	−314.5	−315.0	−314.0
Pasoh, Malaysia	−363.8	−365.3	−363.7
Lambir, Malaysia	−390.1	−391.2	−390.5
Sinharaja, Sri Lanka	−258.8	−258.5	−258.9

A comparison is shown between the analysis by Chave *et al.*² of the dispersal-limitation model (L_1), the density-dependent symmetric model (L_2) and the dispersal-limitation model proposed in ref. 3 (L_3) (the smaller the absolute value of L , the maximum log-likelihood value, the better the quality of the fit). Both dispersal-limitation models produce almost identical fits and are statistically comparable to the density-dependent model.

- Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. *Nature* **438**, 658–661 (2005).
- Chave, J., Alonso, D. & Etienne, R. S. *Nature* **441**, doi:10.1038/nature04826 (2006).
- Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. *Nature* **424**, 1035–1037 (2003).

doi:10.1038/nature04827

REVIEWS

The search for signs of recovery of the ozone layer

Elizabeth C. Weatherhead¹ & Signe Bech Andersen²

Evidence of mid-latitude ozone depletion and proof that the Antarctic ozone hole was caused by humans spurred policy makers from the late 1980s onwards to ratify the Montreal Protocol and subsequent treaties, legislating for reduced production of ozone-depleting substances. The case of anthropogenic ozone loss has often been cited since as a success story of international agreements in the regulation of environmental pollution. Although recent data suggest that total column ozone abundances have at least not decreased over the past eight years for most of the world, it is still uncertain whether this improvement is actually attributable to the observed decline in the amount of ozone-depleting substances in the Earth's atmosphere. The high natural variability in ozone abundances, due in part to the solar cycle as well as changes in transport and temperature, could override the relatively small changes expected from the recent decrease in ozone-depleting substances. Whatever the benefits of the Montreal agreement, recovery of ozone is likely to occur in a different atmospheric environment, with changes expected in atmospheric transport, temperature and important trace gases. It is therefore unlikely that ozone will stabilize at levels observed before 1980, when a decline in ozone concentrations was first observed.

The ozone layer protects all living organisms from excess ultraviolet radiation. Ozone is continually produced, destroyed and circulated in the Earth's atmosphere by a variety of natural influences^{1–4}. In the early 1970s, Molina and Rowland⁵ first theorized that chlorine compounds—manufactured as refrigerants and aerosol propellants—could destroy ozone in the Earth's stratosphere, a theory confirmed in 1985 by a measured ozone loss over Antarctica^{6,7}. Stratospheric ozone levels significantly declined until the mid-1990s⁸ with the largest rates of depletion near the poles and no noticeable depletion near the Equator.

Anthropogenic ozone loss spurred policy makers to ratify a series of international agreements—the Montreal Protocol and its amendments—which, beginning in 1989, led to reduced production and use of ozone-depleting substances⁹. Figure 1a highlights the success of the Protocol by tracing effective equivalent stratospheric chlorine (EESC), a combined measure of stratospheric chlorine and bromine. EESC peaked in the mid- to late-1990s and has started to decline in the past few years^{10,11}. Today, controls on ozone-depleting substances often focus on returning EESC to the level that prevailed in 1980. The turnaround in EESC suggests that the early signs of an ozone layer recovery may be detectable in the early part of this century^{12,13}.

Although previous work has examined changes in ozone trends at 40 km altitude^{14,15}, ozone amounts at this altitude contribute very little to total column amounts, thus the observed changes at 40 km have little effect on ultraviolet radiation levels and their associated human health and environmental effects. The first studies examining total column ozone are now emerging¹⁶. In light of the importance of the ozone layer to the biosphere, we need to verify our understanding of ozone loss and confirm the effectiveness of the Montreal Protocol. We ask whether the observed changes in ozone abundances are consistent with our current understanding of atmospheric processes. It is vital to ask not only whether ozone concentrations are levelling off or increasing, but whether these changes are occurring where and when we are expecting them.

The recovery of the ozone layer is a process beginning with a lessening in the rate of decline, followed by a levelling off and an eventual increase in ozone driven by changes in the concentrations of ozone-depleting substances. The analyses of ozone records and conclusions regarding recovery, in the short or long term, are sensitive to many concurrent changes in the atmosphere. Because of high natural variability in ozone levels, total column ozone fluctuates over timescales of a few years. These fluctuations can obscure long-term changes and offer false indications of recovery. The separation of long-term changes in ozone concentrations from natural variability is our current challenge. Even when ozone-depleting substances are significantly diminished, other anthropogenic changes to the atmosphere will further complicate the recovery process and may result in considerably altered ozone levels in the future. The detection of a thickening of the ozone layer can only be evaluated if naturally occurring processes and events are appropriately accounted for.

In this Review, ozone trends are presented as a function of latitude, season and altitude, to determine if the patterns of change indicate recovery. Observed ozone data are compared with a variety of models to help determine if both the magnitude and patterns of change are what can be expected due to declining levels of ozone-depleting substances. To predict future ozone levels, we examine model estimates for ozone in the year 2050, when the levels of chlorine and bromine in the stratosphere should have returned to near pre-1980 values.

Estimating past and future ozone levels

Atmospheric models (Box 1) are used to estimate both past and future ozone levels and they need to take into account atmospheric chemistry, dynamics and temperature, solar inputs and volcanic eruptions, all of which affect ozone amounts. Figure 1b shows 14 two- and three-dimensional (2D and 3D) model estimates of past and future total column ozone levels for 60° S–60° N (refs 17–30). The models show general agreement with respect to ozone depletion and recovery; however, the rates of depletion and recovery can vary

¹Cooperative Institute for Research in Environmental Science, Campus Box 216, University of Colorado, Boulder, Colorado 80307, USA. ²Danish Meteorological Institute, Lyngbyvej 100, DK-2100, Copenhagen, Denmark.

by a factor of three. The three models that include the effect of the solar cycle in their estimates of future ozone show different estimates of the magnitude of the solar cycle.

The model-calculated abundances of column ozone are normalized to 1980 levels to remove differences between measured and modelled column ozone as high as 30 Dobson units (total column ozone is measured in Dobson units (DU), the number of ozone molecules in a column of atmosphere; at standard pressure and temperature, 330 DU would be 3.3 mm thick). Because of the normalization to 1980, when solar activity was at a peak, this representation overestimates the difference in depletion between measurements and models for 60°N–60°S. However, this bias is removed when trends are derived and presented below. Only about half of the models predict that column ozone will rise above 1980 levels when the abundance of ozone-depleting substances returns to 1980 levels. These differences in recovery rates from the various models result from differences in assumptions about transport, temperature and trace gas concentrations. Examining where and when the models agree with the observed data provides insight about recent changes and whether they can be described as recovery of the ozone layer due to a decline in ozone-depleting substances¹⁷.

Factors affecting ozone levels

Understanding how chemical processes, atmospheric dynamics, temperatures, solar activity and volcanic eruptions affect ozone in the long and short term, and incorporating these factors into models, helps bring model estimates into agreement with observed data. However, many of the effects of these factors are not fully understood, which limits our ability to attribute ozone changes to particular factors. Effects of some of the most important factors and their interactions, together with the limitations in our understanding, are outlined below.

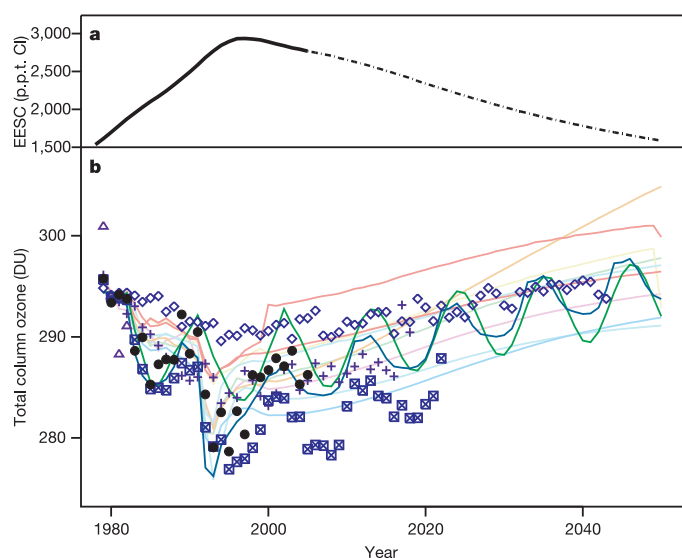


Figure 1 | Reductions in stratospheric chlorine levels compared to total ozone through time. **a**, Measurements of effective equivalent stratospheric chlorine (EESC, in parts per trillion; solid black line) through to 2005 indicate that the Montreal Protocol and its amendments are succeeding in reducing atmospheric chlorine loading. The dashed line represents projected stratospheric chlorine loading through to 2050 based on compliance with the existing agreements. Highest EESC is estimated to have occurred in 1996–98. **b**, Two-dimensional (solid lines) and three-dimensional (all symbols except filled circles) model estimates of past and future total column ozone levels for 60°S–60°N (Box 1). Filled circles are measured ozone amounts from satellites. The models show general agreement with respect to ozone depletion and recovery when averaged over this large region.

Although much of the chemistry affecting ozone is understood, uncertainties in chemical reaction rates, along with the difficulty in estimating future levels of trace gases in the atmosphere, limit our ability to predict recovery of ozone concentrations. For example, uncertainties in bromine chemistry limit our ability to predict recovery of ozone concentrations³¹, and the rates of ozone loss by two key catalytic cycles involving anthropogenic halogens are at present uncertain by about a factor of two³². Changes in stratospheric water and methane are difficult to predict yet will affect total ozone levels^{33,34}, especially if anthropogenic influences alter their concentrations^{35–37}.

Atmospheric dynamics, especially transport of ozone-rich air

Box 1 | Data analysis

All analyses were carried out using advanced statistical approaches with the assumption of autoregressive components in the noise with time lag one month (AR-1). For details see, for instance, ref. 16. The data were also analysed using the CUSUM approach¹³; results are similar with a high dependence on the impact of the solar cycle. The analyses have been carried out on surface Dobson data, with trend results in notable agreement: small differences in declines between data sets were in agreement with small differences in increases.

To understand ozone trends for near-background aerosol conditions, the 24 consecutive months of data that show the strongest effects due to the Mt Pinatubo eruption have been removed from both model output and the measured data. The exact two years are determined separately for each model or data set. The omitted two-year period differs by location owing to the observed aerosol transport times to high latitudes. For the most part, the results are not highly sensitive to small changes in the starting or ending dates of the omitted data.

Trends reported in Figs 3–5 are based on a contiguous, piecewise linear fit to the data using both quasi-biennial oscillation and solar proxies to estimate the effects of these parameters. Uncertainty estimates were computed for Figs 3 and 5 assuming an AR-1 noise term. The statistical confidence levels are strongest for the decline in ozone. Estimates were calculated for the change in trend from the 1979–95 decline and for indications of a statistically significant upward trend. The results show that the change in trend is outside the 2 σ limits for mid-latitudes and high latitudes, but the increase in trends is only significant north of 50°N. Both the derived trends and uncertainty estimates north of 50°N are strongly influenced by the very low ozone values observed in the mid- and late-1990s at these latitude bands. Inclusion of solar proxies, Arctic Oscillation, North Atlantic Oscillation and quasi-biennial oscillation have a little impact on trends but have a substantial impact on the derived error bars on the trends. For additional details, see refs 16 and 17. Data before 1979 are not included in trend analyses presented here.

Ten 2D and four 3D models were used in this analysis. All of the modelling groups were represented in the WMO Ozone Assessment¹¹, but when more recent runs of the model were available, those were used. The 2D models represented are RIVM, AER, ULAQ, GSFC-INT, GSFC, NOCAR, SUNY, OSLO, UIUC and MPI. The 3D models used are GSFC, UMETRAC, NIES and CMAM. Further details on these models are available elsewhere^{11,17}.

EESC showed a near linear increase, which began to level off in approximately 1996. For this reason, many studies look for changes in the rate of ozone decline starting at that date^{14–16}, but changing the date used for the analysis would not significantly affect the results¹⁶. Although an ozone turnaround could be expected to occur at the same time as the peak in EESC loading, the timing of the start of the turnaround in ozone resulting from decreases in ozone-depleting substances is difficult to identify because a number of other factors influence ozone amounts. Examination of the model output offers an even wider range of dates than assumed in these studies. The models, driven by surface concentrations of halogens, suggest that the date when total column ozone may start increasing is confounded by the effects of the Mt Pinatubo eruption and the turnaround could have started before 1996 or might not begin until after 2010.

poleward and strong polar vortices, have direct influences on ozone levels in the short term^{38–41}. Climate change could have long-term effects on atmospheric dynamics throughout the atmosphere. Neither dynamical processes nor dynamical effects on ozone levels are fully understood. For example, recent dynamical activity could be due to natural variability, climate change, or radiative feedbacks resulting from changing concentrations of ozone. Increases in tropopause heights are associated with lower ozone amounts^{42–44}, whereas changes in both the Arctic and North Atlantic oscillations since the early 1990s are associated with higher levels of ozone^{16,45,46}. Climate change may affect dynamical processes by increasing the strength of polar vortices, thereby enhancing polar depletion of ozone^{47–49}, or it could lead to increases in the magnitude of planetary waves causing slightly higher ozone levels⁵⁰.

Ozone levels are directly affected by the temperature of the stratosphere, which is influenced by the stability of the wintertime polar vortex circulation^{38,48,51}. Temperature declines of the past two decades in the lower stratosphere are consistent with expectations of rising greenhouse gas concentrations and observed changes in stratospheric water and ozone amounts^{11,52,53}. In the upper stratosphere, colder conditions shift the balance of ozone photochemistry towards higher ozone concentrations⁵⁴. At high latitudes, colder conditions in the lower stratosphere promote the formation of polar stratospheric clouds which contribute to severe ozone depletion, a condition that occurred several times in the Arctic in the 1990s⁴⁸ and as recently as the spring of 2005. Conversely, in four of the last six years, warmer conditions and less stable Arctic vortices have led to higher levels of ozone in the Arctic⁵¹.

Solar variations directly influence both the destruction and production of ozone⁵⁵. Eleven-year solar cycles can obscure trends, particularly when ozone changes are examined using data over a period of less than two solar cycles. For example, during the 1999–2003 solar maximum, the increases in solar activity may have increased total column ozone levels. However, some short-term solar proton events can cause localized ozone depletions of 30–60% in the upper stratosphere⁵⁶, and energetic particles can cause a downward descent and catalytic destruction of ozone, particularly above the middle stratosphere^{57–59}.

Major volcanic eruptions inject sulphur into the stratosphere, forming sulphate aerosols that react to destroy ozone^{60–63}. Major eruptions of El Chichón in 1982 and Mt Pinatubo in 1991 both occurred during solar maxima, yet had their own large effects on ozone for several years following the events⁶⁴. Despite the spread of Mt Pinatubo aerosols in the stratosphere of both hemispheres, the eruption resulted in a sharp decline in Northern Hemisphere ozone levels^{60–62,65,66} but not in Southern Hemisphere levels¹¹.

Signatures of recovery of the ozone layer

Just as with climate change, where fingerprints of changes are helpful for attribution⁶⁷, signatures of changes in ozone help ensure that observations can be correctly interpreted in terms of the removal of ozone-depleting substances from the atmosphere. Progress in detecting and attributing ozone recovery is most likely to be realized by looking for expected changes when and where they will be optimally detected. Previous work^{11,68} and recent analysis¹⁷ support three signatures as a basis for investigating ozone recovery. These signatures correspond to the latitudinal, seasonal and altitudinal dependences of observed changes. For each of these three identified signatures, we compare the trends derived from the past eight years of data with past depletion rates and model estimates of recovery. Agreement in these three signatures of the magnitude of observed trends will help provide evidence that the ozone layer is recovering. However, analyses on such a short timescale will probably result in large uncertainties because of natural variability. Temperature, winds and other factors have varying effects on stratospheric ozone abundances: the effects can be better understood through detailed analyses of recent atmospheric conditions, which provide further insight

into the causes of changes in ozone concentrations⁶⁹. The focus here is to determine if the recent ozone trends are in agreement with what is expected from a levelling off and decrease in atmospheric chlorine and bromine levels, rather than being influenced mainly by natural variability. One strength of this study is the use of a variety of models, which can offer insight into the current range of expected recovery rates. No one test can prove attribution, yet each can offer additional insight and evidence for likely causes of change.

Figure 2 presents a sample of the monthly averaged time series of total column ozone from 1979 through to the end of 2005 from a merged TOMS/SBUV(/2) satellite data set (version 8). In the last eight years of each of the time series a departure from the downward trend observed in the first fifteen years of the data set is apparent at some latitudes, but there is considerable variability that cannot be attributed to concentrations of ozone-depleting substances. Standard statistical techniques are used to remove from the ozone data the effects of variability due to solar inputs and dynamical factors^{8,16}. One problem with this approach is that different environmental parameters can behave in a similar manner, making it difficult to separate influences through analytical techniques. Another problem is that, in some cases, ozone levels, stratospheric circulation, and temperatures interact and limit attempts to attribute changes through a purely analytical approach. The scientific difficulty concerns interpretation of the recent data—as a sign of ozone recovery, as simple variability, or as a combination of several parameters that affect ozone. Despite the existing uncertainties, latitudinal, seasonal and altitudinal signatures of recovery stand out as robust in the model projections because they spatially and temporally link current ozone increases to patterns of post-1980 ozone depletion.

Latitudinal signature of recovery. Past ozone depletion, shown in red in Fig. 3, has been near zero at the Equator, while gradually increasing towards the poles. The model projections through to 1995 qualitatively reproduce this feature, although most of the projections do not estimate the magnitude of depletion observed, particularly for high latitudes. Ozone trends derived from data for 1996–2004 are

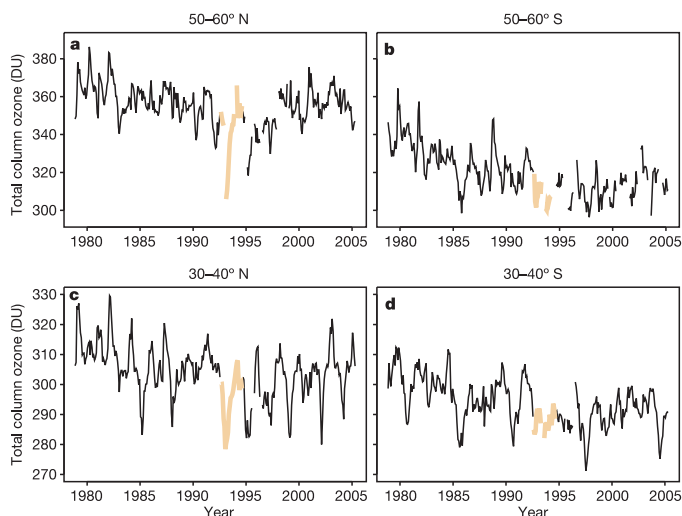


Figure 2 | Deseasonalized ozone data from satellite data at four latitudes. **a**, 50–60° N; **b**, 50–60° S; **c**, 30–40° N; **d**, 30–40° S. The time series illustrate the apparent change in trend that has occurred at mid-latitudes (**c**, **d**). The data removed from the analyses because of possible effects of the Mt Pinatubo eruption are marked in orange. The data contain influences of both the quasi-biennial oscillation and the solar cycle. The changes are most notable in the Northern Hemisphere, particularly for the high latitudes (**a**, **b**), but are not in line with expected changes owing to the current loading of ozone-depleting substances. Data are from version 8 of the TOMS/SBUV(/2) data set.

shown in blue and contrast with the long-term ozone recovery predicted from models. Long-term ozone recovery rates from models are estimated from available time slices for the 3D models and from 1996 to 2050 for 2D models. Again, both the data and models show increases but the amount of the increase observed in the high northern latitudes is considerably larger than what the models predict. This region also exhibits the highest level of natural variability. In the Antarctic, the ozone layer continues to reach severe, low levels in the spring, with some layers in the atmosphere showing nearly complete destruction of ozone^{9,68}. In the Arctic, the situation is more irregular because severe ozone depletion occurs during springtime in years when temperatures are low enough to result in conditions conducive to rapid ozone depletion^{51,70}.

Seasonal signature of recovery. Past ozone depletion has had a distinct seasonal signature in Northern Hemisphere mid-latitudes, with higher rates of ozone depletion observed in the springtime, a feature that the models generally reproduce (Fig. 4). The seasonal aspect of recovery suggested by the recent data is in rough agreement with the modelled patterns, although the recovery is more rapid for most spring and summer months. Similar to the latitudinal signature, the Northern Hemisphere spring is recognized as the season of highest variability, allowing for the possibility of unusual trends being observed from evaluation of short-term data sets. Seasonal trends are not independent, and results have shown that summer ozone levels are strongly linked to spring ozone levels⁷¹. Polar ozone depletion has been strongest during local spring, with influence to the mid-latitudes dependent on the mixing of air and regeneration of ozone⁷².

Altitudinal signature of recovery. Long-term ground-based observations at three stations (Fig. 5) indicate that past ozone depletion

has occurred preferentially in the lower stratosphere. The observed altitudinal dependence is consistent with our understanding of stratospheric photochemistry⁷³. Depletion above 30 km has been small but significant^{74–76}. Ozone concentrations at these higher altitudes, although not contributing significantly to the total column ozone levels, are dominated by changes in chemistry. A turnaround in the decline observed at these layers has been observed^{14,15}. For these three locations, trends in total column ozone are a result of the balance between both positive and negative trends at different layers. At both Arosa and Boulder, the changes in total column ozone are dominated by changes in the lower stratosphere that are larger than expected from the models, while data from Tatenno are in better agreement with the models. This is the region where the largest recovery rates are expected, but it is also the region of greatest natural variability⁷⁷.

Predicting trends in ozone levels

The most severe ozone depletion has been observed in the polar regions, but detecting recovery near the poles will be difficult. In the Antarctic, where conditions are generally conducive to rapid, heterogeneous ozone depletion, ozone at some altitudes can be nearly entirely depleted during the spring. In recent years, total column ozone values have been observed to reach comparably low levels during the Antarctic spring. This apparent levelling off is not attributable to a decline in chlorine loading, but rather is a sign of near-complete depletion at critical layers in the atmosphere. Little improvement is expected for total column ozone in the Antarctic for the next several decades. In contrast, increases in total column ozone in the Arctic will partially depend on possible dynamical and temperature changes in the coming decades, which could result in either an expedited or delayed ozone increase^{47,48,50,78}. Changes in ozone concentrations at specific altitudes in both the Arctic and Antarctic will be highly dependent on temperature and resulting

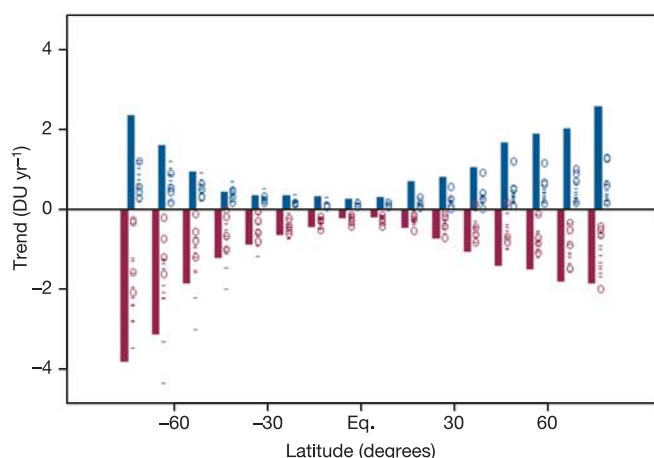


Figure 3 | Measured and modelled ozone trends by latitude for 1979–95 (in red) and 1996–2005 (in blue). Increases in ozone during the past ten years coincide with the regions in which past depletion has occurred.

Analyses of the measurements are represented by the solid bars; analyses of 2D models are shown with horizontal dashes, and results of 3D models are shown with circles. The model estimates for 1979–95 are shown in red; projections for long-term recovery are shown in blue. Trends before 1995 are statistically significant (2σ) for mid and high latitudes. Trends after 1995 show a statistically significant change from the observed decline rates, but the positive trends are not, in general, statistically significant—that is, the derived estimates for the last ten years are consistent with level ozone amounts within estimates of statistical uncertainty. Note that the models and measurements agree well in the Southern Hemisphere, and in the Northern Hemisphere show rough agreement for past trends but significant disagreement for the magnitude of emerging trends. The largest natural variability is observed away from the Equator, allowing for potentially spurious trends in these regions. Data are from version 8 of the merged TOMS/SBUV2 data set. Eq., Equator; negative latitudes are °S.

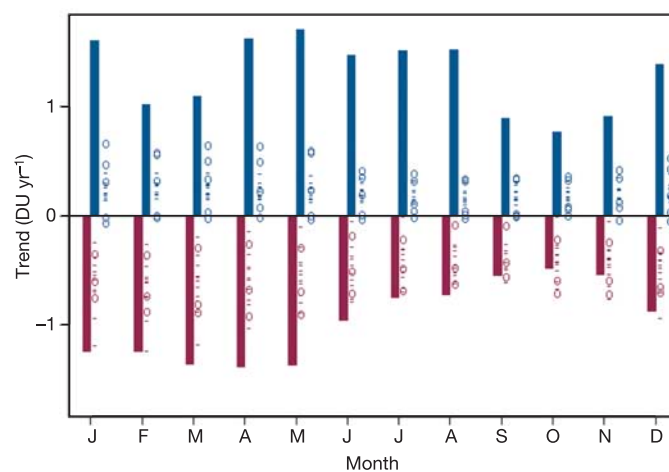


Figure 4 | Measured and modelled seasonal ozone trends at 35° N for 1979–95 (in red) and 1996–2005 (in blue). Recent increases in ozone have occurred at roughly the time of year when past depletion has taken place. The solid bars represent the trends in the data, and the horizontal dashes represent the results from 11 2D models. The open circles represent the five 3D models, each with two years of data removed to avoid the major effects of Mt Pinatubo. The model estimates for 1979–95 are shown in red; long-term projections are shown in blue. Trends before 1995 are statistically significant (2σ) for all seasons. Trends after 1995 show a statistically significant change from the observed decline rates, but the positive trends are not statistically significant for this latitude band—that is, the derived estimates for the last ten years are consistent with level ozone amounts within estimates of statistical uncertainty. Note that the trends over the past ten years show considerably higher rates of increase than expected due to a decrease in EESS. The data are from version 8 of the merged TOMS/SBUV2 data set.

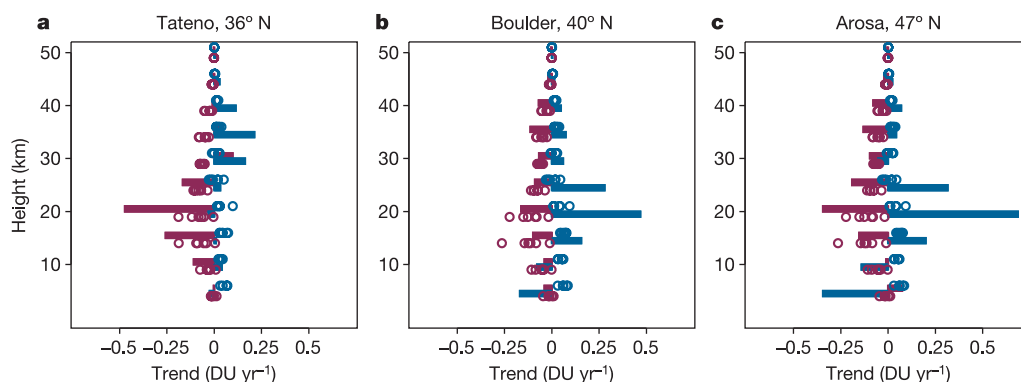


Figure 5 | Measured and modelled ozone trends by altitude at three surface monitoring stations for 1979–95 (in red) and 1996–2005 (in blue). The large positive trends in the most recent data from the lower stratosphere—particularly for Boulder (b) and Arosa (c), and less so for Tateno (a)—are much larger than being predicted by the models for recovery. Trends before 1995 are statistically significant (2σ) for the lower and upper stratosphere,

although not for the mid-stratosphere. Trends after 1995 show a statistically significant change from the observed decline rates, with significant positive trends in the upper stratosphere. This is also a region of the atmosphere characterized by high natural variability and strongly influenced by temperature and dynamics. Data are from the revised Umkehr algorithm⁷⁷.

polar stratospheric cloud formation, as well as on the amounts of ozone-depleting substances.

Complicating the interpretation of recent ozone data are the effects of natural variability, particularly temperature and dynamical influences as well as solar activity. The most striking data illustrating the effects of temperature on the ozone layer are observations from the northern high latitudes, particularly north of 45°N (Fig. 6). Examination of the latitudinal, seasonal and altitudinal signatures of recovery shows that there is coarse agreement between the models and measurements, except in the north, where the data imply much faster recovery than is indicated by the models. The time series illustrate that the trends are strongly affected by the low ozone levels occurring in the mid- to late-1990s, the timeframe in which we begin to look for signs of recovery. During this time not only was EESC peaking, but the Arctic stratosphere was extremely cold and ozone levels were anomalously low⁷⁸. Most of the subsequent years have been warm (with a notable exception of the winter of 2004/05), and correspond to an apparent change in the ozone trend⁵¹. These unusually low ozone years in the mid- to late-1990s affect the derived trends before 1996, but also after 1996 because the record for the most recent eight years starts from an unusually depleted level. Thus, the representation of the latest increases in ozone north of 40°N as evidence of a long-term upward trend and recovery due to declining levels of halogens might be misleading. A change to colder conditions in the Arctic stratosphere would probably lead to reversed trends, at least in the short term. Whereas the effects of these cold winters are well studied for the Arctic, data shown in Fig. 6 suggest that the very low ozone values extend to mid-latitudes as well. The short data record combined with the impact of unusually low ozone levels in the mid- to late-1990s make it difficult to estimate the magnitude of the long-term rates of change in ozone.

The solar cycle strongly influences the amount of ozone and complicates the interpretation of data. Figure 1b shows that all three models that include the 11-year solar cycle estimate an increase in ozone for the past eight years. This is a short-term change, not recovery. The magnitude of the effect of the solar cycle differs from model to model, and may be larger than is indicated by the data. Efforts to derive the size of the influence of the solar cycle from past column ozone time series are complicated by the fact that the prior two solar maxima nearly coincided with the last two major volcanic eruptions. These eruptions offset the increase in ozone expected from increases in solar ultraviolet output. Consequently, some of the increase of total column ozone during the past eight years may be due to the recent solar cycle maximum, but the magnitude of this influence is somewhat uncertain. The data collected over the next few

years will be important for separating the influence of the solar cycle from the long-term recovery of the ozone layer.

On the basis of the analysis of the trends with respect to the latitudinal, seasonal and altitudinal signatures and with respect to the known mechanisms governing ozone levels, at least some of the observed changes in ozone depletion rates are in agreement with expected ozone recovery rates. The rates of change, particularly the increases north of 45°N , are significantly larger than would be initially expected from the small decrease in the concentrations of ozone-depleting substances in the atmosphere. At least two major parameters have contributed to these higher-than-expected ozone levels: the solar cycle, which peaked in 2000–02; and the combination of cold stratospheric temperatures observed over the past ten years and planetary wave driving observed in the northern polar region. Both of these parameters resulted in lower ozone levels in the mid- to late-1990s and higher ozone levels in most of the subsequent years.

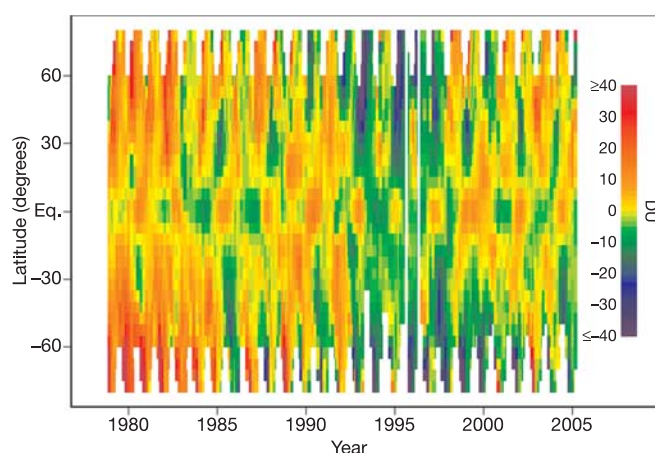


Figure 6 | Deseasonalized total column ozone by latitude. Data are from merged satellite data sets. The very low ozone levels observed in the Arctic in 1996, 1997 and 1998 were well studied and linked to the extremely cold stratospheric temperatures observed in those winters. Because the timing of these very cold winters is close to the turnaround in ozone-depleting substances, the ozone data seem to show a strong change in direction starting in the mid- to late-1990s. It is likely that if the Arctic stratosphere is cold in the near future, ozone will be similarly low. Note that the unusually low ozone conditions appear to extend as far south as 40°N . Colour scale shows deviations of measured total column ozone (in DU) from seasonally expected monthly averages for each latitude band.

Long-term trends are not expected from solar or volcanic activity but variability in ozone as a result of each complicates recovery detection. Ozone data from the very late 1990s and early 2000s coincide with a period of low volcanic activity favouring higher abundances of ozone. Should another major eruption occur in the next few years, ozone levels are likely to be lower, and, depending on the magnitude of the eruption, they could be similar to those observed in the 1990s.

Expectations for ozone levels near the end of this century

A question often asked is 'will ozone return to pre-1980 levels, and if so, when?' Because many of the factors influencing ozone levels are also changing, even if all anthropogenic ozone-depleting substances were removed from the atmosphere, ozone levels might not stabilize at pre-1980 levels^{11,34}. Total column ozone, carbon dioxide emissions, stratospheric temperatures and circulation patterns are closely linked, and changes in one of these variables can affect the others^{11,79–82}. By the end of the century, provided the concentrations of ozone-depleting substances decrease, ozone levels are expected to be dominated by temperature, atmospheric dynamics and the abundances of trace gases, including water vapour, methane and N₂O. For example, future growth in N₂O, due in part to increased fertilizer production, could lead to decreases in ozone. Some model calculations indicate that ozone could increase to a higher level than that observed before the influences of ozone-depleting substances, while other models indicate that ozone could increase, but reach lower levels. Clarifying the expectations for ozone amounts near the end of this century will require improved estimates of future impacts of the various factors as well as continued improvements in the models to represent the combined effects on ozone of these processes.

Remaining uncertainties

Although many factors affect ozone concentrations and it is difficult to be confident about trend results derived from ten years of data, many of the changes observed in the recent ozone data are qualitatively consistent with what would be expected on the basis of a reduction in the concentrations of ozone-depleting substances in the atmosphere. Over the past ten years, total column ozone values for most of the world have levelled off or show a slight increase. No area shows significant depletion of total ozone, marking the first ten-year period since 1980 (omitting the perturbation following the eruption of Mt Pinatubo) in which ozone has not declined. The observed levelling off of ozone is generally consistent with declines in ozone-depleting substances due to international agreements controlling their production. However, the large increases in column ozone observed in the mid- to high-northern latitudes are probably due to the higher temperatures in the Arctic polar vortex as well as to solar influence and the recent lack of volcanic activity. In contrast, the winter of 2004/05 was extremely cold in the Arctic stratosphere, allowing for severe ozone depletion before the polar vortex broke up. The 11-year solar cycle recently peaked and its influence on ozone is uncertain, but solar variability appears to have contributed to some of the observed levelling off and increases in ozone over the past eight years.

As concentrations of ozone-depleting substances subside, considerable uncertainty about the rate of ozone recovery and future ozone levels exist. In the future, ozone levels will depend on continued compliance with the Montreal Protocol and its amendments and on climate change policies that influence atmospheric changes. Changes in the atmosphere as a result of continued anthropogenic impacts suggest that ozone will recover in an atmosphere much different from that which prevailed before the build-up of ozone-depleting substances. Whether ozone stabilizes at a level higher or lower than pre-1980 levels, the vertical distribution of ozone in the future is almost certain to be different from the pre-depletion period. Because recovery approaching pre-1980 levels could take decades, the amount of ultraviolet radiation reaching the Earth's surface is likely to remain elevated for as long as ozone

remains below historically normal values. Data through to the end of this decade are needed to help determine how much of the recent ozone increases are attributable to solar influences, and to establish the extent to which temperatures and concentrations of ozone-depleting substances are affecting current ozone amounts. During the next few years, ozone levels in the Arctic will be strongly influenced by stratospheric temperature, possibly resulting in delayed recovery or record-low ozone observations. Considerably longer data series and improved understanding of atmospheric processes and their effects on ozone are needed to estimate future ozone levels with confidence.

1. Crutzen, P. J. Ozone production rates in an oxygen-hydrogen-nitrogen atmosphere. *J. Geophys. Res.* **76**, 7311–7327 (1971).
2. Johnston, H. S. & Graham, R. Photochemistry of HO_x and HNO₃ compounds. *Can. J. Chem.* **52**, 1415–1423 (1974).
3. Stolarski, R. S. & Cicerone, R. J. Stratospheric chlorine: a possible sink for ozone. *Can. J. Chem.* **52**, 1610–1615 (1974).
4. Wofsy, S. C., McElroy, M. B. & Yung, Y. L. The chemistry of atmospheric bromine. *Geophys. Res. Lett.* **2**, 215–218 (1975).
5. Molina, M. J. & Rowland, F. S. Stratospheric sink for chlorofluoromethanes: Chlorine atom catalysed destruction of ozone. *Nature* **249**, 810–812 (1974).
6. Farman, J., Gardiner, B. & Shanklin, J. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature* **315**, 207–210 (1985).
7. Solomon, S., Garcia, R. R., Rowland, F. S. & Wuebbles, D. J. On the depletion of Antarctic ozone. *Nature* **321**, 755–758 (1986).
8. Staehelin, J., Harris, N. R. P., Appenzeller, C. & Eberhard, J. Ozone trends: A review. *Rev. Geophys.* **39**, 231–290 (2001).
9. Solomon, S. Stratospheric ozone depletion: a review of concepts and history. *Rev. Geophys.* **37**, 275–316 (1999).
10. Montzka, S. A. *et al.* Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature* **398**, 690–694 (1999).
11. World Meteorological Organization. *Scientific Assessment of Ozone Depletion: 2002* (<http://ozone.unep.org/Publications/index.asp>) (2003).
12. Weatherhead, E. C. *et al.* Detecting the recovery of total column ozone. *J. Geophys. Res.* **105**, 22201–22210 (2000).
13. Reinsel, G. C. *et al.* On detection of turnaround and recovery in trend for ozone. *J. Geophys. Res.* **107**, 4078, doi:10.1029/2001JD000500 (2002).
14. Reinsel, G. C. Trend analysis of upper stratospheric Umkehr ozone data for evidence of turnaround. *Geophys. Res. Lett.* **29**, 1451, doi:10.1029/2002GL014716 (2002).
15. Newchurch, M. J. *et al.* Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery. *J. Geophys. Res.* **108**, 4507, doi:10.1029/2003JD003471 (2003).
16. Reinsel, G. C. *et al.* Trend analysis of total ozone data for turnaround and dynamical contributions. *J. Geophys. Res.* **110**, D16306, doi:10.1029/2004JD004666 (2005).
17. Andersen, S. B. *et al.* Comparison of recent modeled and observed trends in total column ozone. *J. Geophys. Res.* **111**, D02303, doi:10.1029/2005JD006091 (2006).
18. Velders, G. J. M. *Scenario Study of the Effects of CFC, HCFC, and HFC Emissions on Stratospheric Ozone* (RIVM Report 722201006, National Institute of Public Health and the Environment, Bilthoven, The Netherlands, 1995).
19. Rinsland, C. P. *et al.* Post-Mount Pinatubo eruption ground-based infrared stratospheric column measurements of HNO₃, NO, and NO₂ and their comparison with model calculations. *J. Geophys. Res.* **108**, 4437, doi:10.1029/2002JD002965 (2003).
20. Pitari, G. & Rizi, V. An estimate of the chemical and radiative perturbation of stratospheric ozone following the eruption of Mt. Pinatubo. *J. Atmos. Sci.* **50**, 3260–3276 (1993).
21. Rosenfield, J. E., Douglass, A. R. & Considine, D. B. The impact of increasing carbon dioxide on ozone recovery. *J. Geophys. Res.* **107**, 4049, doi:10.1029/2001JD000824 (2002).
22. Fleming, E. L. *et al.* Simulation of stratospheric tracers using an improved empirically based two-dimensional model transport formulation. *J. Geophys. Res.* **104**, 23911–23934 (1999).
23. Portmann, R. W. *et al.* Role of nitrogen oxides in the stratosphere: A re-evaluation based on laboratory data. *Geophys. Res. Lett.* **26**, 2387–2390 (1999).
24. Weisenstein, D. K. *et al.* The effects of sulphur emissions from HSCT aircraft: A 2-D model intercomparison. *J. Geophys. Res.* **103**, 1527–1547 (1998).
25. Smyshlyaev, S. P. *et al.* A two-dimensional model with input parameters from a general circulation model: ozone sensitivity to different formulations for the longitudinal temperature variation. *J. Geophys. Res.* **103**, 28373–28387 (1998).
26. Stordal, F., Isaksen, I. S. A. & Hornbrevth, K. A diabatic circulation two-dimensional model with photo-chemistry: Simulations of ozone and long-lived tracers with surface sources. *J. Geophys. Res.* **90**, 5757–5776 (1985).
27. Grobß, J. U., Brühl, C. & Peter, T. Impact of aircraft emissions on tropospheric and stratospheric ozone, I, Chemistry and 2-D model results. *Atmos. Environ.* **32**, 3173–3184 (1998).

28. Wuebbles, D. J. *et al.* New methodology for ozone depletion potentials of short-lived compounds: n-propyl bromide as an example. *J. Geophys. Res.* **106**, 14551–14571 (2001).
29. Schnadt, C. *et al.* Interaction of atmospheric chemistry and climate and its impact on stratospheric ozone. *Clim. Dyn.* **18**, 501–517 (2002).
30. Austin, J. A three-dimensional coupled chemistry-climate model simulation of past stratospheric trends. *J. Atmos. Sci.* **59**, 218–232 (2002).
31. Salawitch, R. J. *et al.* Sensitivity of ozone to bromine in the lower stratosphere. *Geophys. Res. Lett.* **32**, L05811, doi:10.1029/2004GL021504 (2005).
32. Sander, S. P. *et al.* *Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies* Evaluation No. 14, JPL Publ. No. 02-25 (Jet Propulsion Laboratory, Pasadena, California, 2003).
33. Kirk-Davidoff, D. B. *et al.* The effect of climate change on ozone depletion through changes in stratospheric water vapour. *Nature* **402**, 399–401 (1999).
34. Randeniya, L. K., Vohralik, P. F. & Plumb, I. C. Stratospheric ozone depletion at northern midlatitudes in the 21st century: the importance of future concentrations of greenhouse gases nitrous oxide and methane. *Geophys. Res. Lett.* **29**, 1051, doi:10.1029/2001GL014295 (2002).
35. Oltmans, S. J. *et al.* The increase in stratospheric water vapor from balloonborne frostpoint hygrometer measurements at Washington, D.C., and Boulder, Colorado. *Geophys. Res. Lett.* **27**, 3453–3456 (2000).
36. Shindell, D. T. Climate and ozone response to increased stratospheric water vapor. *Geophys. Res. Lett.* **28**, 1551–1554 (2001).
37. Randel, W. J. *et al.* Interannual changes of stratospheric water vapor and correlations with tropical tropopause temperatures. *J. Atmos. Sci.* **61**, 2133–2148 (2004).
38. Chipperfield, M. P. & Jones, R. L. Relative influences of atmospheric chemistry and transport on Arctic ozone trends. *Nature* **400**, 551–554 (1999).
39. Salby, M. L. & Callaghan, P. F. Fluctuations of total ozone and their relationship to stratospheric air motions. *J. Geophys. Res.* **98**, 2715–2727 (1993).
40. Hood, L. L., Rossi, S. & Beulen, M. Trends in lower stratospheric zonal winds, Rossby wave breaking behaviour, and column ozone at northern midlatitudes. *J. Geophys. Res.* **104**, 24321–24339 (1999).
41. Fusco, A. C. & Salby, M. L. Interannual variations of total ozone and their relationship to variations of planetary wave activity. *J. Clim.* **12**, 1619–1629 (1999).
42. Steinbrecht, W., Claude, H., Kohler, U. & Hoinka, K. P. Correlations between tropopause height and total ozone: implications for long-term changes. *J. Geophys. Res.* **103**, 19183–19192 (1998).
43. Forster, P. M. & Tourpali, K. Effect of tropopause height changes on the calculation of ozone trends and their radiative forcing. *J. Geophys. Res.* **106**, 12241–12251 (2001).
44. Santer, B. D. *et al.* Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science* **301**, 479–483, doi:10.1126.1084123 (2003).
45. Zhou, S. *et al.* Trends of NAO and AO and their associations with stratospheric processes. *Geophys. Res. Lett.* **28**, 4107–4110 (2001).
46. Appenzeller, C., Weiss, A. K. & Staehelin, J. North Atlantic Oscillation modulates total ozone winter trends. *Geophys. Res. Lett.* **27**, 1131–1134 (2000).
47. Austin, J. *et al.* Uncertainties and assessments of chemistry-climate models of the stratosphere. *Atmos. Chem. Phys.* **3**, 1–27 (2003).
48. Shindell, D. T., Rind, D. & Loneragan, P. Increased polar stratospheric ozone losses and delayed eventual recovery owing to increasing greenhouse-gas concentrations. *Nature* **392**, 589–592 (1998).
49. Newman, P. A. & Nash, E. R. Quantifying the wave driving of the stratosphere. *J. Geophys. Res.* **105**, 12485–12497 (2000).
50. Schnadt, C. *et al.* Interaction of atmospheric chemistry and climate and its impact on stratospheric ozone. *Clim. Dyn.* **18**, 501–517 (2002).
51. Manney, G. *et al.* The remarkable 2003–2004 winter and other recent warm winters in the Arctic stratosphere since the late 1990s. *J. Geophys. Res.* **110**, D4107, doi:10.1029/2004JD005367 (2005).
52. Ramaswamy, V. Stratospheric temperature trends: observations and model simulations. *Rev. Geophys.* **39**, 71–122 (2001).
53. Shine, K. P. *et al.* A comparison of model simulated trends in stratospheric temperatures. *Q. J. R. Meteorol. Soc.* **129**, 1565–1588 (2003).
54. Thrush, B. A. The chemistry of the stratosphere. *Rep. Prog. Phys.* **51**, 1341–1371 (1988).
55. Zerefos, C. S. *et al.* Solar activity-total ozone relationships: observations and model studies with the heterogeneous chemistry. *J. Geophys. Res.* **102**, 1561–1569 (1997).
56. Randall, C. E. *et al.* Stratospheric effects of energetic particle precipitation in 2003–2004. *Geophys. Res. Lett.* **32**, L05802, doi:10.1029/2004GL022003 (2005).
57. Jackman, C. H., Fleming, E. L. & Vitt, F. M. Influence of extremely large solar proton events in a changing stratosphere. *J. Geophys. Res.* **105**, 11659–11670 (2000).
58. Stephenson, J. A. E. & Scourfield, M. W. J. Importance of energetic solar protons in ozone depletion. *Nature* **352**, 137–139 (1991).
59. Sinnhuber, M. *et al.* A model study of the impact of magnetic field structure on atmospheric composition during solar proton events. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017265 (2003).
60. Dessler, A. E. *et al.* Balloon-borne measurements of ClO, NO, and O₃ in a dense volcanic cloud: an analysis of heterogeneous chemistry between 20 and 30 km. *Geophys. Res. Lett.* **20**, 2527–2530 (1993).
61. Fahey, D. W. *et al.* In situ measurements constraining the role of sulphate aerosols in mid-latitude ozone depletion. *Nature* **363**, 509–514 (1993).
62. Solomon, S. *et al.* The role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes. *J. Geophys. Res.* **101**, 6713–6727 (1996).
63. Prather, M. J. Catastrophic loss of stratospheric ozone in dense volcanic clouds. *J. Geophys. Res.* **97**, 10187–10191 (1992).
64. Thomason, L. W., Poole, L. R. & Deshler, T. A global climatology of stratospheric aerosol surface area density deduced from SAGE II measurements. *J. Geophys. Res.* **102**, 8967–8976 (1997).
65. Schoeberl, M. R., Bhartia, P. K. & Herman, J. R. Tropical ozone loss following the eruption of Mt. Pinatubo. *Geophys. Res. Lett.* **20**, 29–32 (1993).
66. Hofmann, D. J. *et al.* Ozone loss in the lower stratosphere over the United States in 1992–1993: Evidence for heterogeneous chemistry on the Pinatubo aerosol. *Geophys. Res. Lett.* **21**, 65–68 (1994).
67. Hegerl, G. C. *et al.* Multi-fingerprint detection and attribution analysis of greenhouse gas-, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.* **13**, 613–634 (1997).
68. Hofmann, D. J. *et al.* Ten years of ozonesonde measurements at the south pole: implications for recovery of springtime Antarctic ozone. *J. Geophys. Res.* **102**, 8931–8943 (1997).
69. Hadjinicolaou, P., Pyle, J. A. & Harris, N. R. P. The recent turnaround in stratospheric ozone over northern middle latitudes: A dynamical modeling perspective. *Geophys. Res. Lett.* **32**, L12821, doi:10.1029/2005GL022476 (2005).
70. Rex, M. *et al.* Arctic ozone loss and climate change. *Geophys. Res. Lett.* **31**, L04116, doi:10.1029/2003GL018844 (2004).
71. Fioletov, V. E. & Shepherd, T. G. Seasonal persistence of midlatitudes total ozone anomalies. *Geophys. Res. Lett.* **30**, 1417, doi:10.1029/2002GL016739 (2003).
72. Prather, M. & Jaffe, A. H. Global impact of the Antarctic ozone hole: chemical propagation. *J. Geophys. Res.* **95**, 3473–3492 (1990).
73. Logan, J. A. *et al.* Trends in the vertical distribution of ozone: a comparison of two analyses of ozonesonde data. *J. Geophys. Res.* **104**, 26373–26399 (1999).
74. Randel, W. J. *et al.* Trends in the vertical distribution of ozone. *Science* **285**, 1689–1692 (1999).
75. Bojkov, R. D. *et al.* Vertical ozone distribution characteristics deduced from ~44,000 re-evaluated Umkehr profiles (1957–2000). *Meteorol. Atmos. Phys.* **79**, 127–158 (2002).
76. Miller, A. J. *et al.* Comparisons of observed ozone trends and solar effects in the stratosphere through examination of ground-based Umkehr and combined solar backscattered ultraviolet (SBUV) and SBUV 2 satellite data. *J. Geophys. Res.* **101**, 9017–9022 (1996).
77. Petropavlovskikh, I. *et al.* On shifts in the long-term Umkehr radiance records and their influence on retrieved ozone profiles. *Geophys. Res. Lett.* **28**, 255–258 (2001).
78. Waugh, D. W. *et al.* Persistence of the lower stratospheric polar vortices. *J. Geophys. Res.* **104**, 27191–27201 (1999).
79. Pawson, S. & Naujokat, B. The cold winters of the middle 1990s in the northern lower stratosphere. *J. Geophys. Res.* **104**, 14209–14222 (1999).
80. Tabazadeh, A. & Cordero, E. New directions: stratospheric ozone recovery in a changing atmosphere. *Atmos. Environ.* **38**, 647–649 (2004).
81. Rosenfield, J., Douglass, A. R. & Considine, D. B. The impact of increasing carbon dioxide on ozone recovery. *J. Geophys. Res.* **107**, 4049, doi:10.1029/2001JD000824 (2003).
82. Langematz, U. *et al.* Thermal and dynamical changes of the stratosphere since 1979 and their link to ozone and CO₂ changes. *J. Geophys. Res.* **108**, 4027, doi:10.1029/2002JD002069 (2003).

Acknowledgements We thank NASA GSFC, NOAA ESRL, EPA CISES, Danish National Science Foundation, EU CANDIDOZ and the Fulbright Foundation for their support of this research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to E.C.W. (betsy.weatherhead@cires.colorado.edu).

ARTICLES

Shugoshin collaborates with protein phosphatase 2A to protect cohesin

Tomoya S. Kitajima¹, Takeshi Sakuno^{1,3}, Kei-ichiro Ishiguro¹, Shun-ichiro Iemura⁴, Tohru Natsume⁴, Shigehiro A. Kawashima^{1,2} & Yoshinori Watanabe^{1,2,3}

Sister chromatid cohesion, mediated by a complex called cohesin, is crucial—particularly at centromeres—for proper chromosome segregation in mitosis and meiosis. In animal mitotic cells, phosphorylation of cohesin promotes its dissociation from chromosomes, but centromeric cohesin is protected by shugoshin until kinetochores are properly captured by the spindle microtubules. However, the mechanism of shugoshin-dependent protection of cohesin is unknown. Here we find a specific subtype of serine/threonine protein phosphatase 2A (PP2A) associating with human shugoshin. PP2A colocalizes with shugoshin at centromeres and is required for centromeric protection. Purified shugoshin complex has an ability to reverse the phosphorylation of cohesin *in vitro*, suggesting that dephosphorylation of cohesin is the mechanism of protection at centromeres. Meiotic shugoshin of fission yeast also associates with PP2A, with both proteins collaboratively protecting Rec8-containing cohesin at centromeres. Thus, we have revealed a conserved mechanism of centromeric protection of eukaryotic chromosomes in mitosis and meiosis.

Sister chromatid cohesion is carried out by a multisubunit complex, cohesin, consisting of two SMC (structural maintenance of chromosome) family proteins—a kleisin subunit Scc1/Rad21 and an accessory subunit Scc3 (also called stromal antigen (SA) in animal cells)^{1–3}. Cohesion is maintained until metaphase when sister kinetochores attach to microtubules emanating from the opposite spindle poles. The cohesion at centromeres is especially important at this stage, because the establishment of bipolar spindle attachment depends on the tension generated by the pulling force of spindle microtubules and the counteracting force of centromeric cohesion of sister chromatids⁴. Indeed, in animal mitotic cells centromeric cohesin (and cohesion) persists until metaphase, whereas most cohesin dissociates from chromosome arms during prophase and prometaphase to resolve sister chromatids³. At the onset of anaphase, the anaphase promoting complex (APC)-dependent degradation of securin allows the activation of a specific endopeptidase, separase, which in turn cleaves and cleans off residual chromosomal Scc1/Rad21, allowing the separation of sister chromatids⁵. Thus, the dissociation of cohesin is regulated by at least two mechanisms in animal cells. During meiosis, the temporally staggered release of arm and centromeric cohesion is most striking. At the first meiotic division (meiosis I), Rec8—which replaces Scc1/Rad21 in meiosis—is cleaved along chromosome arms but is protected at centromeres, where it is only cleaved during the second division (meiosis II)^{6,7}.

In yeast and probably most eukaryotes, shugoshin (Sgo/MEI-S332) protects meiotic Rec8-containing cohesin from separase cleavage at meiosis I^{6–12}. Human shugoshin (hSgo1; also called shugoshin-like 1 (SGOL1)), which is also expressed during proliferation, protects cohesin at centromeres for mitosis^{13–15}. Phosphorylation of the cohesin subunit SA2 (an Scc3 homologue) by Polo-like kinase 1 (Plk1) is critical for prophase dissociation because the inactivation of Plk1, or the expression of a non-phosphorylatable form of SA2, substantially blocks dissociation of cohesin in early mitosis^{16–18}. Moreover, the dissociation of sister chromatids in hSgo1-depleted

cells is suppressed by expressing this mutant SA2 (ref. 15). In budding yeast and human cells, phosphorylation of the Scc1 subunit by Plk1 enhances its cleavability by separase^{17,19,20}, and this may similarly apply for the meiotic counterpart Rec8 (refs 21, 22). Therefore, a mechanism to protect cohesin at centromeres might be to inhibit its phosphorylation, but no evidence for this has been obtained as yet. It is also unknown whether shugoshin uses a similar mechanism to protect centromeric cohesin in both mitosis and meiosis. Therefore, we have investigated the mechanism by which shugoshin protects cohesin at the centromere.

Shugoshin associates with protein phosphatase 2A

To better understand shugoshin function, we sought to identify associating proteins by tagging hSgo1 with the Flag epitope and expressing the fusion protein in human embryonic kidney (HEK) 293T cells. Anti-Flag immunoprecipitates were analysed using liquid chromatography, followed by tandem mass spectrometry (LC–MS/MS)²³. The majority of peptides identified in the analysis were those of serine/threonine protein phosphatase 2A (PP2A) (Fig. 1a). PP2A is known to act as a heterotrimeric complex consisting of a core dimer of the catalytic C subunit (PP2A-C) and the scaffold A subunit (PP2A-A), which recruits a third variable regulatory B subunit (PP2A-B/PR55/B55, PP2A-B'/PR61/B56, PP2A-B'' or PP2A-B''') that controls substrate specificity or localization of PP2A²⁴. Our MS analysis identified both core subunits PP2A-A and PP2A-C, and most isoforms of the regulatory B56 subunit, but not any isoforms of other B subunits, suggesting that hSgo1 specifically associates with PP2A containing the B56 subunit. Immunoprecipitation analysis of endogenous hSgo1 supported this conclusion (Fig. 1b) and yeast two-hybrid assays suggested direct association of hSgo1 with the PP2A-B56 subunit (Supplementary Fig. 1).

If PP2A functions together with hSgo1, PP2A would localize at the centromere during mitosis. Immunostaining experiments in HeLa cells indicated that the α isoform of PP2A-B56 (PP2A-B56 α)

¹Laboratory of Chromosome Dynamics, Institute of Molecular and Cellular Biosciences, ²Graduate Program in Biophysics and Biochemistry, University of Tokyo, and ³SORST, Japan Science and Technology Agency, Yayoi, Tokyo 113-0032, Japan. ⁴National Institute of Advanced Industrial Science and Technology, Biological Information Research Center, Aomi, Tokyo 135-0064, Japan.

colocalizes with hSgo1 at centromeres from mitotic prophase to metaphase (Fig. 1c). The signals of both proteins decreased at the onset of anaphase. Immunostaining after chromosome spreading further indicated that PP2A-B56 α localizes at the inner centromere between a pair of sister kinetochores (Fig. 1d), like hSgo1 (refs 14, 15, 25). When immunostaining for the core subunits, PP2A-A and PP2A-C, was performed in fixed cells, we found that the signals were dispersed throughout the cell (data not shown). However, by extracting mitotic cells before fixation, we could detect signals of PP2A-A and PP2A-C at centromeres in prometaphase cells but not in anaphase cells (Fig. 1e). Whereas PP2A-B56 α localized only at the inner centromere, PP2A-A and PP2A-C were additionally found at spindle poles during mitosis (Fig. 1e and Supplementary Fig. 2b). These results suggest that the PP2A core complex localizes at various places within mitotic cells, including centromeres, but the PP2A complex containing the B56 subunit preferentially localizes at the inner centromere.

PP2A is required for the protection of centromeric cohesion

To directly assess the importance of PP2A for protecting sister

chromatid cohesion at centromeres, we constructed short interfering (si)RNAs against PP2A-A and treated HeLa cells with them, which resulted in considerable reduction of PP2A-A protein (Fig. 2a). PP2A-A depletion resulted in an accumulation of mitotic cells, with the prometaphase population being particularly increased in number (Fig. 2b). Chromosomes were highly condensed and the number of spindle poles was often increased (Fig. 2c and Supplementary Fig. 2a), consistent with previous observations using okadaic acid, a potent PP2A inhibitor^{26,27}. To examine centromeric cohesion, we spread the chromosomes of mitotic cells treated with PP2A-A siRNAs after incubation with nocodazole for 4 h. In control cells, only ~5% of mitotic cells showed separation of sister chromatids. In PP2A-A-depleted cells, however, ~15% of mitotic cells showed loosened or lost centromere cohesion, and ~30% showed sister chromatid separation (Fig. 2d, e). This separation occurred at prometaphase rather than anaphase, because most PP2A-A-depleted mitotic cells showed positive staining for cyclin B1 (Fig. 2b). Immunostaining of PP2A-A siRNA-treated cells with anti-Rad21 antibodies showed that cohesin localization was accordingly lost in prometaphase cells (Fig. 2f). The poor penetrance of the phenotype

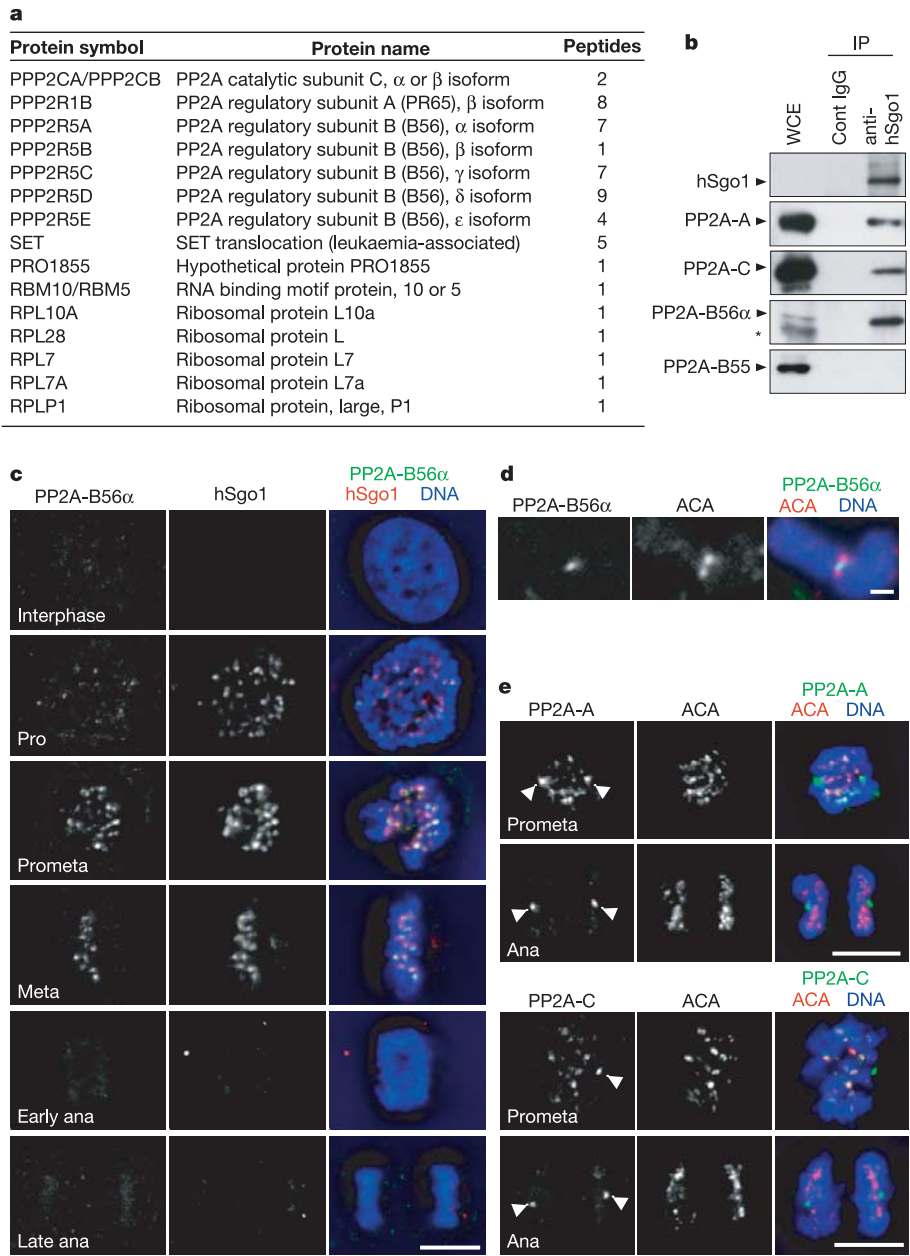


Figure 1 | PP2A associates and colocalizes with hSgo1. **a**, Proteins reproducibly detected in two independent hSgo1 immunoprecipitations by MS analysis are listed. The number of identified peptides of each protein is also shown. **b**, Immunoprecipitates (IP) from an extract of asynchronously growing HeLa cells, obtained using an anti-hSgo1 antibody or control (Cont) IgG, were analysed by western blotting using antibodies against the indicated proteins. An asterisk indicates a cross-reaction. WCE, 0.05% of the whole cell extract. **c**, HeLa cells were stained with anti-PP2A-B56 α (green) and anti-hSgo1 (red) antibodies. DNA was counterstained using Hoechst 33342 (blue). PP2A-B56 α colocalizes with Sgo1 at centromeres from mitotic prophase to metaphase in almost all cells ($n > 50$). Pro, prophase; Prometa, prometaphase; Meta, metaphase; Ana, anaphase. **d**, A single mitotic chromosome stained with anti-PP2A-B56 α (green) and anti-centromere antibodies (ACA) (red). **e**, Mitotic cells were spun onto glass slides using a cytospin and pre-extracted before fixation. The cells were immunostained with anti-PP2A-A or anti-PP2A-C (green) and ACA (red). DNA was stained with Hoechst 33342 (blue). Signals on the spindle poles are indicated by arrowheads. Scale bars, 10 μ m (**c**, **e**) and 1 μ m (**d**).

after exposure to the PP2A-A siRNA can be explained by the residual amount of PP2A-A in the siRNA-treated cells (Fig. 2a), as a more severe phenotype was obtained by treating HeLa cells with okadaic acid (Supplementary Fig. 3). Taken together, we conclude that, like hSgo1, PP2A is required for centromeric protection of sister chromatid cohesion during prophase and prometaphase.

hSgo2 is required for the localization of PP2A at centromeres

Given that hSgo1 associates with PP2A, we examined the possibility that hSgo1 and PP2A require each other for their localization to centromeres (Fig. 3, and see also Supplementary Fig. 4). When PP2A-A was depleted by siRNA, centromeric hSgo1 signals became weakened (Fig. 3a), suggesting that PP2A has a role in facilitating hSgo1 localization to centromeres. In contrast, PP2A-B56 and PP2A-A localization was preserved at centromeres in the hSgo1-depleted cells (Fig. 3c, d), indicating that PP2A can associate with centromeres independently of hSgo1.

Human cells contain another shugoshin-like protein, hSgo2 (also known as SGOL2 and TRIPIN)⁸, which has not been studied as yet. To gain a thorough understanding of the relationship between shugoshin and PP2A, we included hSgo2 in our analysis. We found that hSgo2 localizes at centromeres throughout prophase until metaphase, and disappears at anaphase (Supplementary Fig. 5), which is very similar to the localization of hSgo1 (refs 13–15, 25). Likewise, the depletion of hSgo2 using siRNA caused precocious dissociation of centromeric cohesin and separation of sister chromatids (Fig. 2), indicating that hSgo2 also functions in the centromeric protection of cohesin. The depletion of either hSgo1 or hSgo2 did not influence the localization of the other, indicating that they can independently localize to centromeres (Fig. 3a, b). Notably, the depletion of hSgo2 abolished the localization of PP2A (both the regulatory PP2A-B56 and core PP2A-A subunits) at centromeres (Fig. 3c, d). Consistent with this, PP2A coprecipitates with hSgo2; however, the association may occur through the core subunit PP2A-A

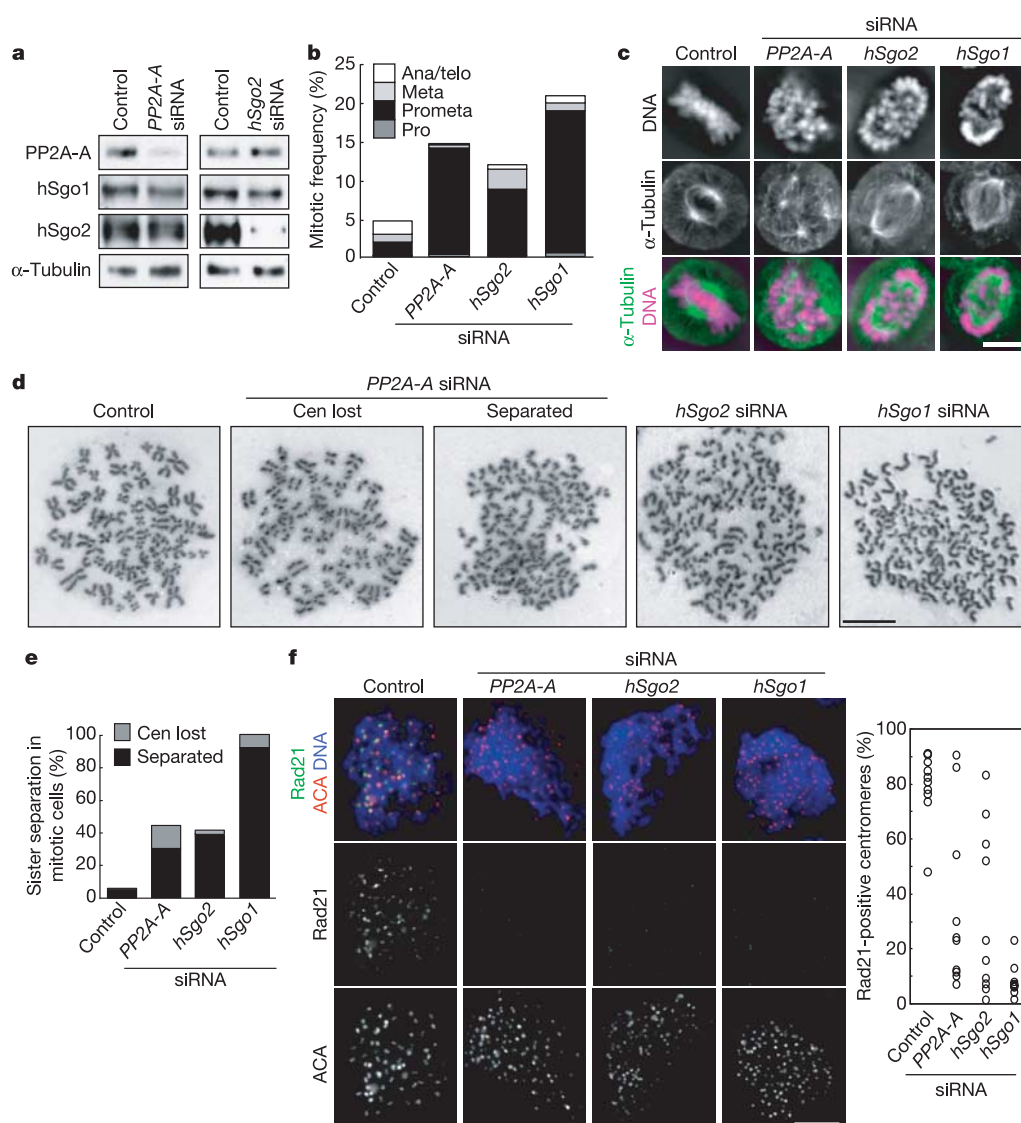


Figure 2 | PP2A is required for centromeric protection. **a**, Extracts from mitotic HeLa cells after exposure to siRNA were immunoblotted with the indicated antibodies. **b**, Mitotic index after siRNA treatment was determined by observing cell shape in living cells ($n > 560$). The mitotic phase was determined by staining for cyclin B1 and DNA in fixed cells ($n > 100$). **c**, Representative prometaphase or metaphase cells stained with anti- α -tubulin (green) and Hoechst 33342 (purple) are shown. **d**, **e**, Mitotic cells after siRNA exposure were treated with nocodazole for 4 h, and

chromosomes were spread and stained with Giemsa (**d**). The frequency of cells showing sister chromatid separation ('Separated') or loss of centromeric cohesion ('Cen lost') was determined ($n > 100$) (**e**). **f**, Mitotic cells treated with the indicated siRNAs were spread by cytopsin and stained with ACA (red) and anti-Rad21 (green). DNA was counterstained with Hoechst 33342 (blue). Average percentage of Rad21-positive centromeres are shown (one dot represents the average of positive centromeres within one cell). Scale bars, 10 μ m (**c**, **d**, **f**).

rather than the regulatory subunit PP2A-B56 (Supplementary Figs 1 and 6). Curiously, when centromeric PP2A was displaced by *hSgo2* siRNA, *hSgo1* localization was not impaired, although it was reduced by the depletion of PP2A (Fig. 3a). This may indicate that *hSgo1* dynamically associates with centromeres, and that cytoplasmic PP2A may influence this interaction. If either *hSgo1* or centromeric PP2A (using *hSgo2* siRNA) is absent from the centromere, the protection of sister centromeres is impaired, suggesting that both proteins collaboratively function at the centromere to protect the dissociation of cohesin.

The shugoshin complex counteracts phosphorylation of SA

Given that phosphorylation of the cohesin subunit SA2 (and presumably SA1) by Plk1 is critical for cohesin dissociation¹⁷, we proposed that SA2 phosphorylation is counteracted by the shugoshin–PP2A complex at centromeres. To examine whether cohesin SA2 preserved around centromeres is indeed dephosphorylated *in vivo*, we prepared nocodazole-treated prometaphase cells in which cohesin is largely dissociated from the chromosome arms but tethered exclusively at centromeres, depending on shugoshin and PP2A (Fig. 2f). The cell extracts were fractionated into chromatin-bound and -unbound fractions, separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), and blotted with anti-SA2 antibodies. As expected, SA2 showed slow electrophoretic migration in the prometaphase chromatin-unbound fraction, representing phosphorylation (presumably by Plk1)¹⁷. In contrast, chromatin-bound SA2 was mostly dephosphorylated (Fig. 4a). These results suggest that SA2 preserved at centromeres is dephosphorylated *in vivo*.

We next examined the biochemical properties of the shugoshin complex *in vitro*. We prepared a carboxy-terminal peptide of SA2

(SA2-C), in which most of the phosphorylation sites are included¹⁷, and phosphorylated it with recombinant Plk1 *in vitro*. The phospho-labelled SA2 fragment was then mixed with purified *hSgo1* complex. The results show that immunoprecipitates of *hSgo1* can dephosphorylate SA2-C, and that this phosphatase activity is inhibitable by okadaic acid (Fig. 4b). Similar phosphatase activity was detected in the *hSgo2* immunoprecipitates. These phosphatase activities also dephosphorylated a C-terminal peptide of SA1 that had been phosphorylated by Plk1, but not histone H3 phosphorylated by Aurora B kinase, indicative of substrate specificity. Thus, the shugoshin complex has the ability to counteract the phosphorylation of Scc3/SA *in vitro*. Taken together, these results support the hypothesis that PP2A-dependent dephosphorylation of Scc3/SA prevents the dissociation of cohesin from centromeres, as part of the centromeric protection function of the shugoshin complex.

PP2A is required for centromeric protection in meiosis

Schizosaccharomyces pombe Sgo1 protects centromeric cohesin containing Rec8 from separase cleavage at meiosis I^{8,9}. In a yeast two-hybrid screen searching for proteins that interact with Sgo1, we frequently isolated Par1—one of the PP2A-B56 homologues in fission yeast²⁸. The interaction between Sgo1 and Par1 was confirmed by immunoprecipitation (Fig. 5a). In proliferating cells, Par1 localizes in the cytoplasm, and on the spindle pole body and the division septum (data not shown)^{28,29}; however, it colocalizes with Sgo1 at centromeres during meiosis I (Fig. 5b). Sgo1 localization was not impaired in mutant *par1Δ* cells (data not shown). In contrast, the centromeric localization of Par1 was abolished in mutant *sgo1Δ* cells (Fig. 5c), indicative of the dependency of Par1 localization on shugoshin and reminiscent of the situation in human mitotic cells.

To examine whether Par1 is required for the protection of centromeric cohesin during meiosis, we analysed Rec8 localization at metaphase II—the period during which Rec8 is detected only at centromeres. We found that, like *sgo1Δ* cells, *par1Δ* cells mostly lost

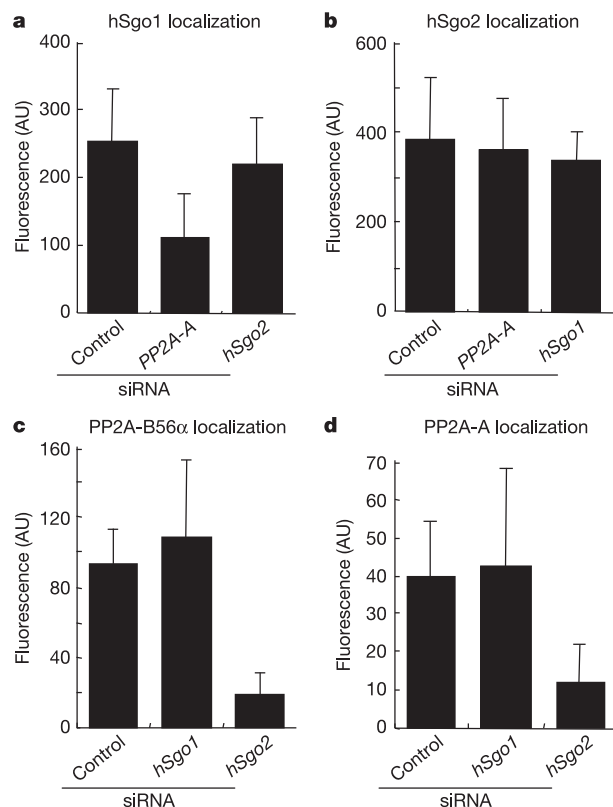


Figure 3 | Interdependency of shugoshin and PP2A for localization. **a–d**, Cells after siRNA treatment were stained with the indicated antibodies (**a**, *hSgo1*; **b**, *hSgo2*; **c**, PP2A-B56; **d**, PP2A-A; see also the representative stained cells in Supplementary Fig. 4). The intensities of the fluorescent centromeric signals in prometaphase cells were quantified as described in Methods. AU, arbitrary units. Error bars represent s.d. ($n = 20$).

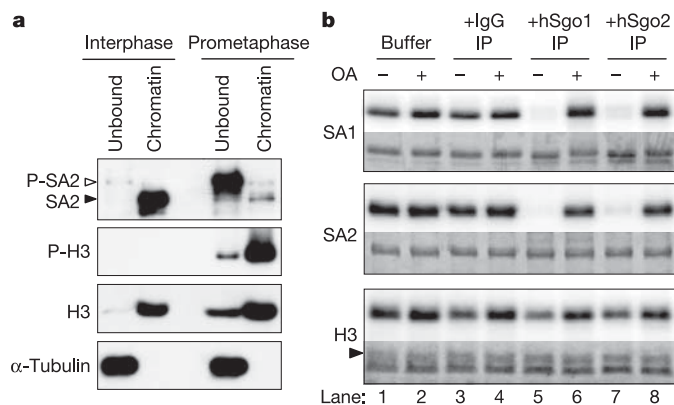


Figure 4 | Dephosphorylation of cohesin subunit SA. **a**, Chromatin-bound SA2 is dephosphorylated *in vivo*. Cell extracts prepared from interphase and prometaphase cells were fractionated into chromatin-bound and -unbound fractions and analysed by western blotting with the indicated antibodies. Four-times more of the chromatin-bound fractions were loaded. P-SA2, phosphorylated SA2; P-H3, histone H3 phosphorylated on Ser 10. In the mitotic chromatin-bound fraction, contaminated interphase chromatin is, if any, negligible (see Supplementary Fig. 9). **b**, Endogenous shugoshin associates with an okadaic-acid-inhibitable phosphatase activity that dephosphorylates cohesin subunit Scc3/SA. C-terminal peptides of SA1 or SA2 phosphorylated by Plk1, and a control histone H3 phosphorylated by Aurora B, were mixed with immunoprecipitation (IP) buffer (lane 1 and 2), control IgG immunoprecipitates (lane 3 and 4), immunoaffinity-purified *hSgo1* (lane 5 and 6) or *hSgo2* (lane 7 and 8) from HeLa cell extracts. Each sample was incubated in the presence (+) or absence (–) of 1 μ M okadaic acid (OA) for 2 h. Autoradiography is shown in the upper panel, Coomassie blue staining in the lower panel.

centromeric Rec8 localization at this stage (Fig. 5d). Consistent with this observation, both of these mutant cell types showed precocious centromeric dissociation after meiosis I, and random chromosome segregation following meiosis II (Supplementary Fig. 7). *S. pombe* cells have another PP2A-B56 homologue, Par2, which is expressed at much lower levels²⁸ and contributes little to centromeric protection in meiosis (data not shown). Taken together, these results argue that, like Sgo1, Sgo1-associating PP2A (including Par1/PP2A-B56) has a crucial role in protecting cohesin at centromeres during meiosis I.

Individual protection ability of PP2A and shugoshin

Our current results in both human cells and fission yeast have raised the possibility that PP2A has an intrinsic ability to protect cohesin. Indeed, by ectopically localizing *S. pombe* Par1/PP2A-B56 to a specific site on a chromosome arm, we observed that the cohesion (and cohesin) at this site was partly preserved even after meiosis I—the period when arm cohesin should dissociate (Supplementary Fig. 8). To assess more thoroughly the individual ability of PP2A for centromeric protection, we used the ‘ectopic protection system’ in fission yeast, where coexpression of Rec8 and its protector blocks sister chromatid separation in mitosis, causing lethality⁸. We

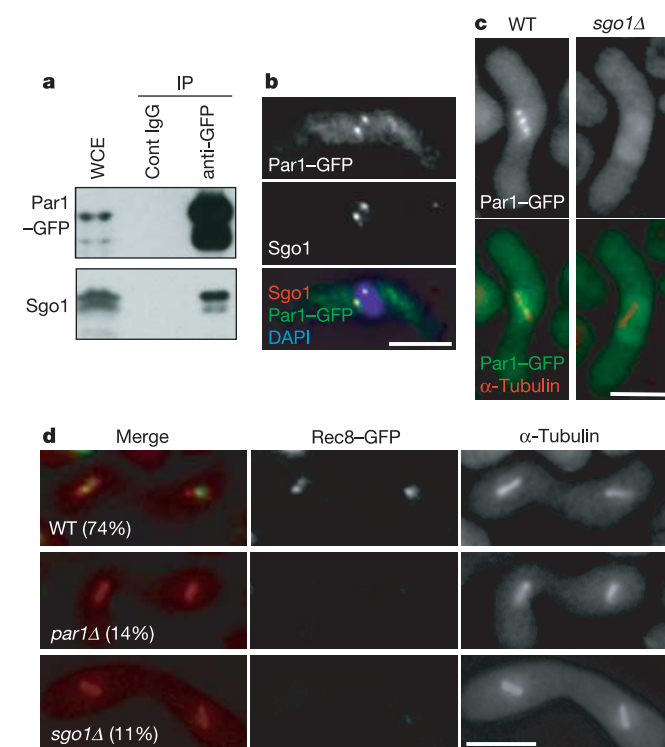


Figure 5 | *S. pombe* PP2A associating with Sgo1 is required for centromeric protection of Rec8-containing cohesin during meiosis I. **a**, Extracts, prepared from mitotic *par1*⁺-GFP cells ectopically expressing Sgo1, were immunoprecipitated with control IgG or anti-GFP (green fluorescent protein) antibodies and analysed by western blotting using antibodies against GFP and Sgo1. WCE, 7.5% of the whole cell extract. **b**, Meiotic *par1*⁺-GFP cells were arrested at metaphase I by repressing APC activation (*slp1*⁺ and *cut23*⁺ expression), and stained with DAPI (4,6-diamidino-2-phenylindole) and antibodies against Sgo1 and GFP. Colocalization was observed in most cells (95%, *n* = 20). **c**, As in **b**, fluorescence of Par1-GFP was examined at metaphase I in wild-type (WT) *sgo1*⁺ and *sgo1*Δ cells. The spindles were visualized by expressing cyan fluorescent protein (CFP)-Atb2 (α2-tubulin). Par1-GFP was detected as dots in most metaphase I *sgo1*⁺ cells (98%, *n* = 219), but never in *sgo1*Δ cells (0%, *n* = 178). **d**, The Rec8-GFP signal was monitored at prometaphase II in the indicated strains. Representative samples are shown together with the frequency of the cells showing centromeric Rec8-GFP (*n* > 50).

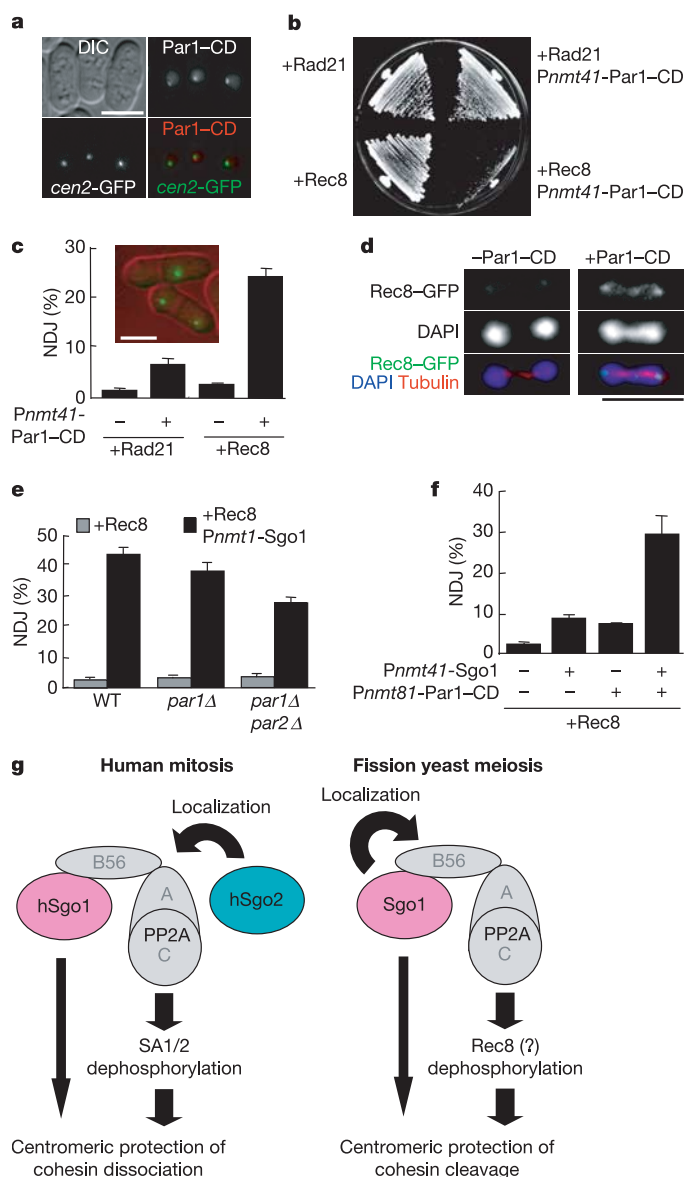


Figure 6 | Individual ability of PP2A and shugoshin for centromeric protection. **a**, Par1-CD proteins visualized by CFP localized in close vicinity to *cen2*-GFP in proliferating cells. **b**, The haploid *cen2*-GFP strains expressing the indicated genes by exogenous promoters (a constitutive promoter, *Padh1*, for *rad21*⁺ and *rec8*⁺, and a thiamine-repressible promoter, *Pnmt41*, for *par1*⁺-CD) were streaked on a thiamine-depleted plate. **c**, The strains in **b** were cultured at 30 °C for 15 h after thiamine depletion and the frequency of NDJ was counted among septated cells. Examples of *cen2*-GFP (green) in *Padh1-rec8*⁺ *Pnmt41-par1*⁺-CD cells are shown. **d**, Centromeric Rec8-GFP signals are detected during anaphase in most *par1*⁺-CD-expressing cells (81%) but in fewer non-expressing cells (19%). **e**, The indicated haploid *cen2*-GFP strains expressing Rec8 or Rec8 and Sgo1 (by the *Pnmt1* promoter) were examined for the frequency of NDJ among septated cells. **f**, The haploid *cen2*-GFP *Padh1-rec8*⁺ strains mildly expressing Sgo1 and/or Par1-CD were examined for the frequency of NDJ. Note the different strength of the thiamine repressible promoters (*Pnmt1* > *Pnmt41* > *Pnmt81*). Error bars represent s.d. of triplicate samples (each *n* > 100) (**c**, **e**, **f**). Scale bars, 5 μm (**a**, **c**, **d**). **g**, Model for the collaboration of shugoshin and PP2A in protecting centromeric cohesin during human mitosis and fission yeast meiosis. PP2A containing the B56 subunit, which is recruited to the centromere by shugoshin (hSgo2 in human cells and Sgo1 in *S. pombe*), dephosphorylates cohesin subunits as a mechanism of protection. Human hSgo1 and *S. pombe* Sgo1 may have an individual activity for centromeric protection, apart from localizing PP2A to centromeres.

proposed that PP2A has an intrinsic ability to protect Rec8 without the help of Sgo1 once it is localized to centromeres. To test this, we endowed Par1 with its own ability to localize to centromeres by fusing its C-terminal end with the chromo domain (CD) of Swi6, which binds to Lys-9-methylated histone H3 largely locating at pericentromeric heterochromatin regions—the sites where Sgo1 usually localizes⁸. The engineered protein, Par1-CD, indeed localized at centromeres in mitotic cells in which Sgo1 is not expressed (Fig. 6a). Notably, coexpression of Par1-CD and Rec8 frequently led to blocked nuclear division, as centromere-associated *cen2*-GFP (green fluorescent protein) frequently segregated to the same side of a septated cell (Fig. 6b, c). This non-disjunction (NDJ) of sister chromatids presumably stems from persistent cohesion at anaphase because centromeric cohesin Rec8 was largely protected at this stage (Fig. 6d). Coexpression of Par1-CD with Rad21 caused a much weaker phenotype, indicative of the specificity of protection for the cohesin kleisin subunit (Fig. 6b, c). Finally, the protection is indeed executed by PP2A activity recruited by Par1-CD because the NDJ was suppressed by introducing a mutation in Ppa2, a major catalytic subunit of PP2A in fission yeast³⁰ (NDJ decreased from 24% to 7%). As Sgo1 is meiosis-specific and is not expressed in mitotic cells, these results demonstrate that centromeric localization of PP2A itself can protect Rec8 cohesin from cleavage, suggesting that dephosphorylation of cohesin Rec8 is a mechanism for the protection of sister centromeres in fission yeast.

We noticed that the ectopic protection mediated by Sgo1 overexpression was alleviated, but not abolished, when endogenous Par1 (or both Par1 and Par2) was depleted from the cells (Fig. 6e). This indicates that Sgo1 also has an individual ability to protect Rec8-containing cohesin without the aid of PP2A-B56 if sufficient amounts of protein are expressed. However, in physiological meiosis, this ability of Sgo1 is not sufficient to complement the loss of PP2A activity. The collaboration of Sgo1 and PP2A was further supported by the observation that coexpression of Sgo1 and Par1-CD mediated a synergistic effect on centromeric protection (Fig. 6f).

Discussion

In animal cells, most cohesin is removed from chromosome arms during prophase and prometaphase, triggered by Plk1-dependent phosphorylation of the Scc3/SA subunit of cohesin¹⁷. Here we have discovered that a B56-containing subtype of PP2A phosphatase associates with hSgo1 in human cells, playing a crucial part in preventing cohesin dissociation at centromeres. We have also demonstrated that chromatin-bound SA2 at centromeres is mostly dephosphorylated in prometaphase cells, whereas dissociated SA2 from the arms is phosphorylated (Fig. 4a). Furthermore, purified hSgo1 or hSgo2 complex can counteract Plk1-dependent SA phosphorylation *in vitro* (Fig. 4b). Thus, our results argue that dephosphorylation of Scc3/SA by shugoshin-associating PP2A is a mechanism for centromeric protection (Fig. 6g). Because dephosphorylated SA2 is detected only in the chromatin-bound fraction of cohesin, we suggest that the regulation takes place exactly at the sites where shugoshin-PP2A complexes localize. Previous results suggested that Plk-dependent phosphorylation of Scc1/Rad21 facilitates its cleavage by separase but is not essential^{17,19,20}, whereas phosphorylation of Rec8 might be crucial for cleavage^{21,22}. Here we have demonstrated that *S. pombe* Par1/PP2A-B56 associating with Sgo1 is required and even sufficient for protecting cohesin Rec8 from separase cleavage. Moreover, PP2A-dependent cohesin protection requires the kleisin subunit Rec8 (Fig. 6b, c), suggesting that the extent of phosphorylation and/or its contribution to the susceptibility to separase cleavage is different between Scc1/Rad21 and Rec8. These results are consistent with the notion that PP2A activity counteracts the phosphorylation of Rec8, thereby protecting it from separase cleavage during meiosis I. Chromosome segregation in mouse meiosis is also disturbed by okadaic acid treatment, which results in premature separation of sister chromatids during meiosis

³¹. Thus, the shugoshin-PP2A system found in *S. pombe* meiosis is presumably applicable to mammalian meiosis. Taken together, centromeric protection of eukaryotic chromosomes may be executed at the level of dephosphorylation of cohesin subunits, and a subtype of PP2A containing the B56 subunit has a direct role in this process. This concept is applicable for both mitosis and meiosis, albeit the crucial target of dephosphorylation is different, being Scc3/SA in mitosis and Rec8 in meiosis (Fig. 6g).

The depletion of hSgo1 by siRNA caused precocious separation of centromeric cohesion (Fig. 2), although the centromeric localization of PP2A is preserved (Fig. 3c, d), suggesting that protection is not solely executed by linking PP2A to centromeres. Moreover, hSgo2 siRNA caused fewer defects in protection than hSgo1 siRNA (Fig. 2e). Therefore, hSgo2 might solely be required to tether PP2A to centromeres, whereas—besides facilitating PP2A function at centromeres—hSgo1 might have an additional role in protection (Fig. 6g). Supporting this notion, ectopic expression of *S. pombe* Sgo1 and Rec8 can enforce centromeric protection even in PP2A-B56-depleted cells (Fig. 6e). Because shugoshin closely associates with cohesin *in vivo*^{8,12,32}, we favour the possibility that Sgo1 physically protects cohesin against access by an inactivating enzyme (for example, Plk1 in human mitosis and Plk or separase in fission yeast meiosis). As Plk is suggested to phosphorylate and delocalize Sgo/MEI-S332 in *Drosophila*³³, PP2A might facilitate the localization of shugoshin, which is supported by our observation that hSgo1 localization partly depends on PP2A (Fig. 3a). Thus, shugoshin and PP2A may support each other, collaboratively protecting cohesin at centromeres. *S. pombe* Sgo1 has roles in both recruiting PP2A and protecting cohesin *per se* at centromeres. In contrast, hSgo1 is dispensable for localizing PP2A to centromeres but is required for centromeric protection, whereas hSgo2 is required for the recruitment of PP2A to centromeres, implying a ‘division of labour’ between these two shugoshin-like proteins in human cells. Thus, the interplay of shugoshin and PP2A is apparently conserved across human and fission yeast, or mitosis and meiosis (Fig. 6g).

PP2A is a family of abundantly expressed protein phosphatases, the activity of which is highly regulated and implicated in a multitude of cellular processes, such as signal transduction, development and tumorigenesis³⁴. Chromosome mis-segregation in mitosis may contribute to tumorigenesis. Meiotic chromosome segregation is also important clinically, as failures in this process cause birth defects in humans. Our study may provide a novel link between PP2A and tumorigenesis or birth defects, and therefore is useful for future studies in those fields as well.

METHODS

Antibody production and immunofluorescence microscopy. Antibody production and immunofluorescence staining were performed as described in Supplementary Methods.

Quantification of fluorescent signals. To quantify the centromeric fluorescent signals, in-focus images of the cyclin B1- or phospho-H3-positive prometaphase cells were taken with the use of MetaMorph imaging software (Universal Imaging). We measured the maximum intensity among the centromeric signals within the cell and subtracted the background intensity of the region, which was measured directly adjacent to the centromere.

RNA interference. Synthetic sense and antisense siRNA oligonucleotides for hSgo1 (ref. 14) and hSgo2 (5'-GCACUACCACUUUGAAUAATT-3'), PP2A- α (5'-AGACUUGACAUUGUUGUUGTT-3') and PP2A- β (5'-UUUCUACUCC AAGUGCUAGTT-3') were obtained from Jbioso. Note that PP2A-B56 consists of five isoforms, whereas the PP2A-A core subunit has only two isoforms. Therefore, we constructed siRNAs against PP2A-A isoforms, rather than PP2A-B56 isoforms, to reduce PP2A activity. The presented data were obtained using the abovementioned siRNAs, but we confirmed that similar results were obtained by using another set of siRNAs: hSgo2 (5'-GCUCUCAUGAACAAUACUTT-3'), PP2A- α (5'-GCAUCAUUGUGUGUGUGATT-3') and PP2A- β (5'-CGAC UCAACAGUAUUAAGATT-3'). Cells at 20% confluency in Opti-MEM medium (Invitrogen) were transfected with siRNA duplexes at a final concentration of 400 nM and Oligofectamine (Invitrogen) at 1:250, and complete medium containing 20% FBS was added at 1:1 after 6 h. After two days incubation, the

cells were examined. All control samples were similarly treated but were exposed to H₂O instead of siRNA reagent.

Preparation of HeLa cell extracts. Preparation of HeLa cell extracts and subsequent immunoprecipitation or fractionation are described in Supplementary Methods.

In vitro dephosphorylation assay. To generate phosphorylated SA substrates, a 6 × His-tagged SA1 C-terminal peptide (amino acids 923–1258) or SA2 C-terminal peptide (amino acids 895–1232) was expressed in *Escherichia coli* strain BL21 and purified with Ni-NTA agarose (Qiagen), and labelled with [γ -³²P]ATP using recombinant Plk1 kinase (ref. 35). As a control, core histone proteins were labelled with [γ -³²P]ATP by recombinant Aurora B kinase. The immunoprecipitated complexes from HeLa cell extracts were collected after washing with immunoprecipitation buffer without phosphatase inhibitors (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1% Tween 20, 10% glycerol and 10 mM β -mercaptoethanol). The equivalent of 3 μ l of immunoprecipitated beads was preincubated with 1 μ M of okadaic acid or dimethylsulphoxide (DMSO) for 20 min at room temperature (~20 °C), followed by the addition of ³²P-labelled SA substrates in a dephosphorylation buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.01% Brij35, 2 mM MnCl₂, 0.1 mM EGTA, 2 mM dithiothreitol (DTT)) supplemented with 1 μ g of BSA to a total reaction volume of 15 μ l and incubated for 2 h at 30 °C with gentle agitation.

Chromosome spreading. Mitotic HeLa cells were collected by mitotic shake-off and treated with 330 nM nocodazole for 4 h. Chromosome spreading was performed as described previously³⁶.

Yeast experiments. All *S. pombe* strains used are listed in Supplementary Table 1. General methods for immunoprecipitation, culturing *S. pombe*, inducing meiosis and monitoring chromosome segregation were as described previously⁸. Further details of *S. pombe* experiments, as well as the yeast two-hybrid assay, are described in Supplementary Methods.

Received 3 October 2005; accepted 21 February 2006.

Published online 15 March 2006.

- Nasmyth, K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* **35**, 673–745 (2001).
- Milutinovich, M. & Koshland, D. E. SMC complexes—wrapped up in controversy. *Science* **300**, 1101–1102 (2003).
- Hirano, T. Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev.* **19**, 1269–1287 (2005).
- Hauf, S. & Watanabe, Y. Kinetochore orientation in mitosis and meiosis. *Cell* **119**, 317–327 (2004).
- Uhlmann, F. Chromosome cohesion and separation: from men and molecules. *Curr. Biol.* **13**, R104–R114 (2003).
- Lee, J. Y. & Orr-Weaver, T. L. The molecular basis of sister-chromatid cohesion. *Annu. Rev. Cell Dev. Biol.* **17**, 753–777 (2001).
- Watanabe, Y. Shugoshin: guardian spirit at the centromere. *Curr. Opin. Cell Biol.* **17**, 590–595 (2005).
- Kitajima, T. S., Kawashima, S. A. & Watanabe, Y. The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. *Nature* **427**, 510–517 (2004).
- Rabitsch, K. P. *et al.* Two fission yeast homologs of *Drosophila* Mei-S332 are required for chromosome segregation during meiosis I and II. *Curr. Biol.* **14**, 287–301 (2004).
- Marston, A. L., Tham, W. H., Shah, H. & Amon, A. A genome-wide screen identifies genes required for centromeric cohesion. *Science* **303**, 1367–1370 (2004).
- Katis, V. L., Galova, M., Rabitsch, K. P., Gregan, J. & Nasmyth, K. Maintenance of cohesin at centromeres after meiosis I in budding yeast requires a kinetochore-associated protein related to MEI-S332. *Curr. Biol.* **14**, 560–572 (2004).
- Hamant, O. *et al.* A REC8-dependent plant shugoshin is required for maintenance of centromeric cohesion during meiosis and has no mitotic functions. *Curr. Biol.* **15**, 948–954 (2005).
- Salic, A., Waters, J. C. & Mitchison, T. J. Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. *Cell* **118**, 567–578 (2004).
- Kitajima, T. S., Hauf, S., Ohsugi, M., Yamamoto, T. & Watanabe, Y. Human Bub1 defines the persistent cohesion site along the mitotic chromosome by affecting Shugoshin localization. *Curr. Biol.* **15**, 353–359 (2005).
- McGuinness, B. E., Hirota, T., Kudo, N. R., Peters, J.-M. & Nasmyth, K. Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. *PLoS Biol.* **3**, e86 (2005).
- Sumara, I. *et al.* The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. *Mol. Cell* **9**, 515–525 (2002).
- Hauf, S. *et al.* Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol.* **3**, e69 (2005).
- Losada, A., Hirano, M. & Hirano, T. Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. *Genes Dev.* **16**, 3004–3016 (2002).
- Alexandru, G., Uhlmann, F., Mechtler, K., Poupart, M. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459–472 (2001).
- Hornig, N. C. & Uhlmann, F. Preferential cleavage of chromatin-bound cohesin after targeted phosphorylation by Polo-like kinase. *EMBO J.* **23**, 3144–3153 (2004).
- Lee, B. H. & Amon, A. Role of Polo-like kinase CDC5 in programming meiosis I chromosome segregation. *Science* **300**, 482–486 (2003).
- Clyne, R. K. *et al.* Polo-like kinase Cdc5 promotes chiasmata formation and cosegregation of sister centromeres at meiosis I. *Nature Cell Biol.* **5**, 480–485 (2003).
- Natsume, T. *et al.* A direct nanoflow liquid chromatography–tandem mass spectrometry system for interaction proteomics. *Anal. Chem.* **74**, 4725–4733 (2002).
- Janssens, V. & Goris, J. Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochem. J.* **353**, 417–439 (2001).
- Tang, Z., Sun, Y., Harley, S. E., Zou, H. & Yu, H. Human Bub1 protects centromeric sister-chromatid cohesion through Shugoshin during mitosis. *Proc. Natl Acad. Sci. USA* **101**, 18012–18017 (2004).
- Vandre, D. D. & Wills, V. L. Inhibition of mitosis by okadaic acid: possible involvement of a protein phosphatase 2A in the transition from metaphase to anaphase. *J. Cell Sci.* **101**, 79–91 (1992).
- Van Dolah, F. M. & Ramsdell, J. S. Okadaic acid inhibits a protein phosphatase activity involved in formation of the mitotic spindle of GH4 rat pituitary cells. *J. Cell. Physiol.* **152**, 190–198 (1992).
- Jiang, W. & Hallberg, R. L. Isolation and characterization of *par1*⁺ and *par2*⁺: two *Schizosaccharomyces pombe* genes encoding B' subunits of protein phosphatase 2A. *Genetics* **154**, 1025–1038 (2000).
- Le Goff, X. *et al.* The protein phosphatase 2A B'-regulatory subunit par1p is implicated in regulation of the *S. pombe* septation initiation network. *FEBS Lett.* **508**, 136–142 (2001).
- Kinoshita, N., Ohkura, H. & Yanagida, M. Distinct, essential roles of type 1 and 2A protein phosphatases in control of the fission yeast cell division cycle. *Cell* **63**, 405–415 (1990).
- Mailhes, J. B., Hilliard, C., Fuseler, J. W. & London, S. N. Okadaic acid, an inhibitor of protein phosphatase 1 and 2A, induces premature separation of sister chromatids during meiosis I and aneuploidy in mouse oocytes *in vitro*. *Chromosome Res.* **11**, 619–631 (2003).
- Kiburz, B. M. *et al.* The core centromere and Sgo1 establish a 50-kb cohesin-protected domain around centromeres during meiosis I. *Genes Dev.* **19**, 3017–3030 (2005).
- Clarke, A. S., Tang, T. T., Ooi, D. L. & Orr-Weaver, T. L. POLO kinase regulates the *Drosophila* centromere cohesion protein MEI-S332. *Dev. Cell* **8**, 53–64 (2005).
- Janssens, V., Goris, J. & Van Hoof, C. PP2A: the expected tumor suppressor. *Curr. Opin. Genet. Dev.* **15**, 34–41 (2005).
- Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E. & Nishida, E. Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *J. Biol. Chem.* **278**, 25277–25280 (2003).
- Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore–microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161**, 281–294 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hauf for critically reading the manuscript. We also thank B. Akiyoshi for yeast two-hybrid screening of *S. pombe* Sgo1; K. Nasmyth for communicating unpublished results; J. M. Peters, E. Nishida, F. Ishikawa, Y. Nakatani, M. Sato and T. Toda for reagents; J. M. Peters, M. Ohsugi and M. Shimura for methods; and all members in our laboratory for their valuable discussion and help. K.I. was supported by a JSPS fellowship. This work was supported in part by the New Energy and Industrial Technology Development Organization (to T.N.), and by the Toray Science Foundation, Uehara Memorial Foundation, and a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to Y.W.).

Author Contributions Experiments in Figs 1–3 were mainly performed by T.S.K., Fig. 4 by K.I., and Figs 5 and 6 by T.S. and S.A.K. Mass spectrometry of hSgo1 immunoprecipitates was performed by S.I. and T.N. Experimental design, interpretation of data, and the preparation of the manuscript were mainly conducted by T.S.K., T.S., K.I. and Y.W.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.W. ([ywatana@iam.u-tokyo.ac.jp](mailto:ywatanab@iam.u-tokyo.ac.jp)).

Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I

Christian G. Riedel^{1*}, Vittorio L. Katis^{1*†}, Yuki Katou², Saori Mori², Takehiko Itoh³, Wolfgang Helmhart^{1†}, Marta Gálová¹, Mark Petronczki¹, Juraj Gregan¹, Bulent Cetin^{1†}, Ingrid Mudrak⁴, Egon Ogris⁴, Karl Mechtler¹, Laurence Pelletier⁵, Frank Buchholz⁵, Katsuhiko Shirahige² & Kim Nasmyth^{1†}

Segregation of homologous maternal and paternal centromeres to opposite poles during meiosis I depends on post-replicative crossing over between homologous non-sister chromatids, which creates chiasmata and therefore bivalent chromosomes. Destruction of sister chromatid cohesion along chromosome arms due to proteolytic cleavage of cohesin's Rec8 subunit by separase resolves chiasmata and thereby triggers the first meiotic division. This produces univalent chromosomes, the chromatids of which are held together by centromeric cohesin that has been protected from separase by shugoshin (Sgo1/MEI-S332) proteins. Here we show in both fission and budding yeast that Sgo1 recruits to centromeres a specific form of protein phosphatase 2A (PP2A). Its inactivation causes loss of centromeric cohesin at anaphase I and random segregation of sister centromeres at the second meiotic division. Artificial recruitment of PP2A to chromosome arms prevents Rec8 phosphorylation and hinders resolution of chiasmata. Our data are consistent with the notion that efficient cleavage of Rec8 requires phosphorylation of cohesin and that this is blocked by PP2A at meiosis I centromeres.

During mitosis, sister chromatids are held together by a multisubunit complex, called cohesin, from their creation during DNA replication until their disjunction at anaphase. Cohesin ensures that sister kinetochores attach to microtubules emanating from opposite spindle poles (known as amphitelic attachment). Cleavage of cohesin's α -kleisin subunit (Scc1/Rad21) by a specialized protease called separase destroys sister chromatid cohesion and triggers the metaphase to anaphase transition¹. Separase is regulated by the binding of an inhibitory chaperone called securin, the destruction of which at the hands of a ubiquitin protein ligase called the anaphase-promoting complex or cyclosome (APC/C) only occurs when all chromosomes have achieved amphitelic attachment².

During meiosis, haploid gametes are formed from diploid precursors by two rounds of chromosome segregation after a single round of DNA replication. The first meiotic division differs from mitosis in two key respects. First, crossing over between homologous non-sister chromatids, which takes place after pre-meiotic DNA replication, creates chiasmata and thereby bivalent chromosomes containing four sets of chromatids. Second, owing to the action of meiosis-specific kinetochore proteins called monopolins^{3,4}, maternal and paternal sister centromere pairs, but not sister centromeres, are pulled in opposite directions by meiosis I microtubules. This creates a novel state of tension, namely between microtubules attempting to pull homologous non-sister centromeres towards opposite poles of the cell and sister chromatid cohesion distal to crossovers that holds bivalents together. This tug of war is resolved by the destruction of

sister chromatid cohesion along chromosome arms through cleavage by separase of cohesin's α -kleisin subunit, in this case a meiosis-specific variant called Rec8 (refs 5, 6). This triggers anaphase I and the first meiotic division. Crucially, cohesin holding sister centromeres together is refractory to cleavage by separase during meiosis I and persists at centromeres until a second round of separase activation at the onset of anaphase II, which finally causes disjunction of sister centromeres, splitting of univalents into individual chromatids, and thereby formation of haploid progeny. Although not necessary for meiosis I, persistence of centromeric sister chromatid cohesion is essential for the amphitelic attachment of sister kinetochores during meiosis II⁵.

Protection of cohesin from separase at meiosis I depends on a class of proteins called shugoshins^{7,8}, which are orthologues of the MEI-S332 protein from *Drosophila melanogaster*⁹. Budding yeast *Saccharomyces cerevisiae*^{10,11} and *Drosophila* possess only a single orthologue (Sgo1 and MEI-S332, respectively) that is expressed in mitotic as well as meiotic cells. The fission yeast *Schizosaccharomyces pombe* possesses two paralogues^{7,8}, one (Sgo1) that is expressed solely during meiosis I and is essential for protecting centromeric sister chromatid cohesion, and a second (Sgo2) that is expressed in both mitotic and meiotic cells. In fission yeast, the association of Sgo1 with centromeres depends on Bub1, a protein that is necessary for delaying activation of the APC/C until all chromosomes have come under tension on metaphase plates as well as for protecting centromeric sister chromatid cohesion¹². Sgo1 also protects centromeric

¹Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria. ²Laboratory of Genome Structure and Function, Division of Gene Research, Center for Biological Resources and Informatics, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan. ³Research Center for Advanced Science and Technology, Mitsubishi Research Institute Inc., 2-3-6 Ohtemachi, Chiyoda-ku, Tokyo 100-8141, Japan. ⁴Department of Medical Biochemistry, Max F. Perutz Laboratories, Vienna Biocenter, Medical University of Vienna, Dr. Bohr-Gasse 9/2, A-1030 Vienna, Austria. ⁵Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Pfotenhauerstrasse, 01307 Dresden, Germany. [†]Present address: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

*These authors contributed equally to this work.

sister chromatid cohesion in mammalian tissue culture cells, where phosphorylation of cohesin's Scc3/SA2 subunit causes cohesin to dissociate from chromosomes via a mechanism (known as the prophase pathway) that does not involve cleavage of α -kleisins by separase^{13,14}. It is not understood how Sgo1 orthologues protect cohesin either from separase during meiosis or from the prophase pathway during mitosis.

Sgo1 binds PP2A at meiosis I

To investigate how Sgo1 protects centromeric cohesin, *S. pombe* and *S. cerevisiae* Sgo1 proteins (called SpSgo1 and ScSgo1, respectively) were purified using a carboxy-terminal tandem affinity purification (TAP) tag¹⁵ from meiosis I cells. Purified material was visualized by SDS–polyacrylamide gel electrophoresis (PAGE) and analysed by tandem mass spectrometry (MS/MS). Notably, for both SpSgo1 and ScSgo1 the most abundant co-purifying proteins were subunits of PP2A (Fig. 1a, b; see also Supplementary Fig. S1a, b and Supplementary Table S1). Importantly, none of these was found associated with 'control' proteins purified in parallel by the very same technique (Fig. 1a, b; see also Supplementary Fig. S1a, b and Supplementary Table S1).

PP2A is mostly found as a hetero-trimeric complex containing catalytic (C), scaffold (A) and regulatory (B, B', B'' or B''') subunits. The *S. pombe* genome encodes two isoforms of the C subunit (SpPpa1 and SpPpa2), which have largely overlapping functions^{16,17}, a single isoform of the A subunit (SpPaa1), one B subunit (SpPab1), and two B' subunits (SpPar1 and SpPar2)^{18,19}. Hereafter, individual subunits of PP2A (or their genes) will be identified as either A, B or C type using the appropriate superscript letter appended to the end of the subunit name. Notably, we detected only a single combination of PP2A subunits associated with SpSgo1, namely SpPaa1^A–SpPar1^{B'}–SpPpa2^C, and an equivalent combination with ScSgo1, namely ScTpd3^A–ScRts1^{B'}–ScPph21^C/ScPph22^C (refs 20, 21), which we refer to hereafter as SpPP2A- π and ScPP2A- π (from the Greek for protector).

To investigate whether the Sgo1–PP2A- π interaction is conserved in mammals, we expressed a functional TAP-tagged mouse Sgo1 protein at physiological levels in HeLa cells²². Mouse Sgo1–TAP, but not a control protein, associated with catalytic, scaffold and all five

isoforms of the B' PP2A- π regulatory subunit of PP2A. Notably, endogenous human Sgo1 also associated with mouse Sgo1–TAP, suggesting that Sgo1 proteins form homo-oligomers (Supplementary Fig. S1a, b and Supplementary Table S1).

The SpSgo1–PP2A- π interaction was confirmed by purifying a TAP-tagged SpPar1^{B'} subunit. Most SpPar1^{B'} was associated with A and C subunits of PP2A, but a minor fraction associated with SpSgo1 (Fig. 1a, b; see also Supplementary Table S1). Interaction of Sgo1 and PP2A- π was reproduced *in vitro*. When synthesized in reticulocyte lysates, SpPar1^{B'} and SpPaa1^A, but not SpPpa2^C, associated with a maltose binding protein (MBP)–SpSgo1 fusion protein immobilized on an affinity matrix but not with MBP alone (Supplementary Fig. S1c). This experiment was repeated using MBP fusions containing different parts of SpSgo1, which revealed that SpPar1^{B'} and SpPaa1^A bound most tightly to SpSgo1's amino-terminal coiled-coil region (Supplementary Fig. S1d).

PP2A- π protects sister chromatid cohesion in *S. pombe*

To test the role of PP2A- π during meiosis, we deleted the genes for SpPpa2^C and SpPar1^{B'} as well as their two closest paralogues in the fission yeast genome, SpPpa1^C and SpPar2^{B'}. Sequences close to *cen1*, marked by a *lac* operator array that binds GFP–LacI (*lys1*–GFP), segregated with high fidelity to all four meiotic products in *Sppa1*^C Δ and *Sppar2*^{B'} Δ cells, mis-segregated modestly in *Spppa2*^C Δ cells, and mis-segregated massively in *Sppar1*^{B'} Δ cells (Supplementary Fig. S2a). Although we did not detect SpPpa1^C associated with SpSgo1 (Fig. 1a, b; see also Supplementary Table S1), it is possible that this isoform can partially replace SpPpa2^C, as in vegetative growth¹⁷.

To analyse chromosome segregation in more detail, we observed segregation of *lys1*–GFP when homozygous in *h*⁹⁰ cells or heterozygous in *h*⁺/*h*[–] cells at anaphase I and II. Deletion of *Sppar1*^{B'} or *Spppa2*^C, like that of *Spsgo1* but not *Spsgo2* (ref. 7), had no effect on meiosis I segregation (Fig. 2a). In contrast, *Sppar1*^{B'} deletion, like that of *Spsgo1*, caused random segregation of sister *lys1*–GFP sequences at meiosis II. *Spppa2*^C deletion had a more modest but nevertheless significant effect (Fig. 2b). Mis-segregation in *S. pombe* is often accompanied by lagging chromosomes during anaphase. Lagging chromosomes were frequently observed at anaphase I in *Spsgo2* mutant cells but rarely in *Spsgo1*, *Spppa2*^C or *Sppar1*^{B'}

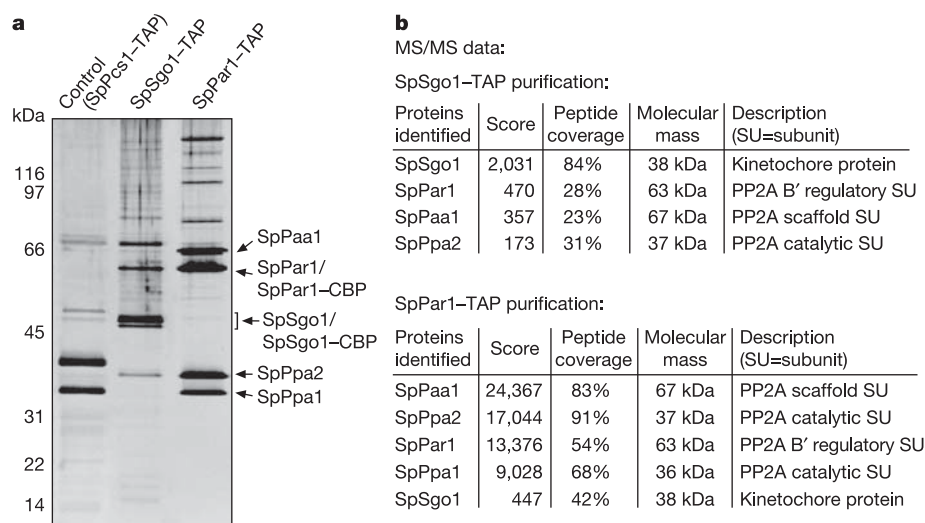


Figure 1 | Sgo1 binds to PP2A- π *in vivo*. **a**, **b**, Pre-starved *S. pombe* *h*[–]/*h*[–] diploids, expressing either TAP-tagged SpSgo1 or SpPar1, were induced to enter synchronous meiosis triggered by Pat1 inactivation⁴⁶ and cells were harvested just before undergoing the first meiotic division. Protein complexes associated with SpSgo1 and SpPar1 were isolated by tandem affinity purification (TAP). Co-purified proteins were separated by SDS–PAGE, visualized by silver-staining (**a**) and identified by tandem mass spectrometry (MS/MS). Proteins that specifically associate with Sgo1 or

Par1 are shown (**b**). The specificity of co-purified proteins was determined by comparison to a 'control' protein SpPcs1, which was isolated by TAP in parallel. Scores shown in the tables were derived from Mascot software. They are measures of the significance of protein identifications. (For more details concerning Mascot or Sequest scores, see Supplementary Information.) All identified proteins shown are highly significant hits. For a full list of co-purified proteins see Supplementary Table S1.

mutants. In contrast, they were frequently observed at anaphase II in *Spsgo1* and *Sppar1^{B'}* mutants but less frequently in *Spppa2^C* or *Spsgo2* mutants (Fig. 2c). These data imply that loss of *SpPar1^{B'}* and *SpSgo1* have similar effects on meiotic chromosome segregation, namely random segregation of sister centromeres at meiosis II after a normal meiosis I. *Sppar1^{B'}* mutants resembled *Spsgo1* mutants in another crucial respect, namely a failure of most cells to retain *SpRec8* at centromeres after meiosis I (Fig. 2d). A similar defect was also observed in a large fraction of *Spppa2^C* mutant cells. In contrast, *SpPP2A- π* mutants had no noticeable effect on the expression, nuclear accumulation or association with centromeres of *SpRec8* before anaphase I (Supplementary Fig. S3a–c).

PP2A- π protection of centromeric cohesin is conserved

To investigate the role of PP2A- π in *S. cerevisiae*, we analysed the effect of eliminating genes encoding either of its two catalytic subunits (*ScPPH21^C* and *ScPPH22^C*) or its single B' regulatory subunit (*ScRTS1^{B'}*) from diploid strains in which the *URA3* locus close to the centromere of chromosome V was marked by a *tet* operator array that binds tetR-GFP (*URA3*-GFP). Homozygous deletion of *ScPPH21^C* or *ScPPH22^C* had little or no effect on meiotic chromosome segregation; however, deletion of *ScRTS1^{B'}* caused massive mis-segregation (Supplementary Fig. S2b). Deletion of

either *ScSGO1* (refs 10, 11) or the spindle checkpoint gene *ScBUB1* (ref. 23) (Supplementary Fig. S2b) caused a greater severity of defects, mainly because these mutations affect meiosis I as well as meiosis II (see Fig. 3c). *ScPP2A- π* can contain either *ScPph21^C* or *ScPph22^C*, and presumably either protein is sufficient to prevent meiosis II mis-segregation. Elimination of both causes a severe growth defect²⁰. A more detailed analysis of chromosome segregation in *Scrts1^{B'}* cells at anaphase I and II (in diploids either homozygous or heterozygous for *URA3*-GFP) revealed that sister centromeres invariably segregate to the same pole at anaphase I but frequently separate precociously (Fig. 3a), as found in *Scsgo1 Δ* (refs 10, 11) and *Schub1 Δ* mutants. This precocious separation is followed by (and presumably causes) high rates of non-disjunction at meiosis II (Fig. 3b). Disjunction of homologous centromeres at meiosis I was normal in the *Scrts1^{B'}* mutant, which contrasts with frequent non-disjunction in *Scsgo1 Δ* (23%) and *Schub1 Δ* cells (46%) (Fig. 3c). This implies that *ScRts1^{B'}* participates in some but not all functions of *ScSgo1*.

To address whether non-disjunction of sister centromeres at meiosis II in *Scrts1^{B'}* cells could be caused by a failure to protect *ScRec8*, we analysed its localization in anaphase I spreads. In wild-type cells, a cluster of centromeric *ScRec8* foci is invariably associated with spindle poles. These clusters were entirely absent in 80% of

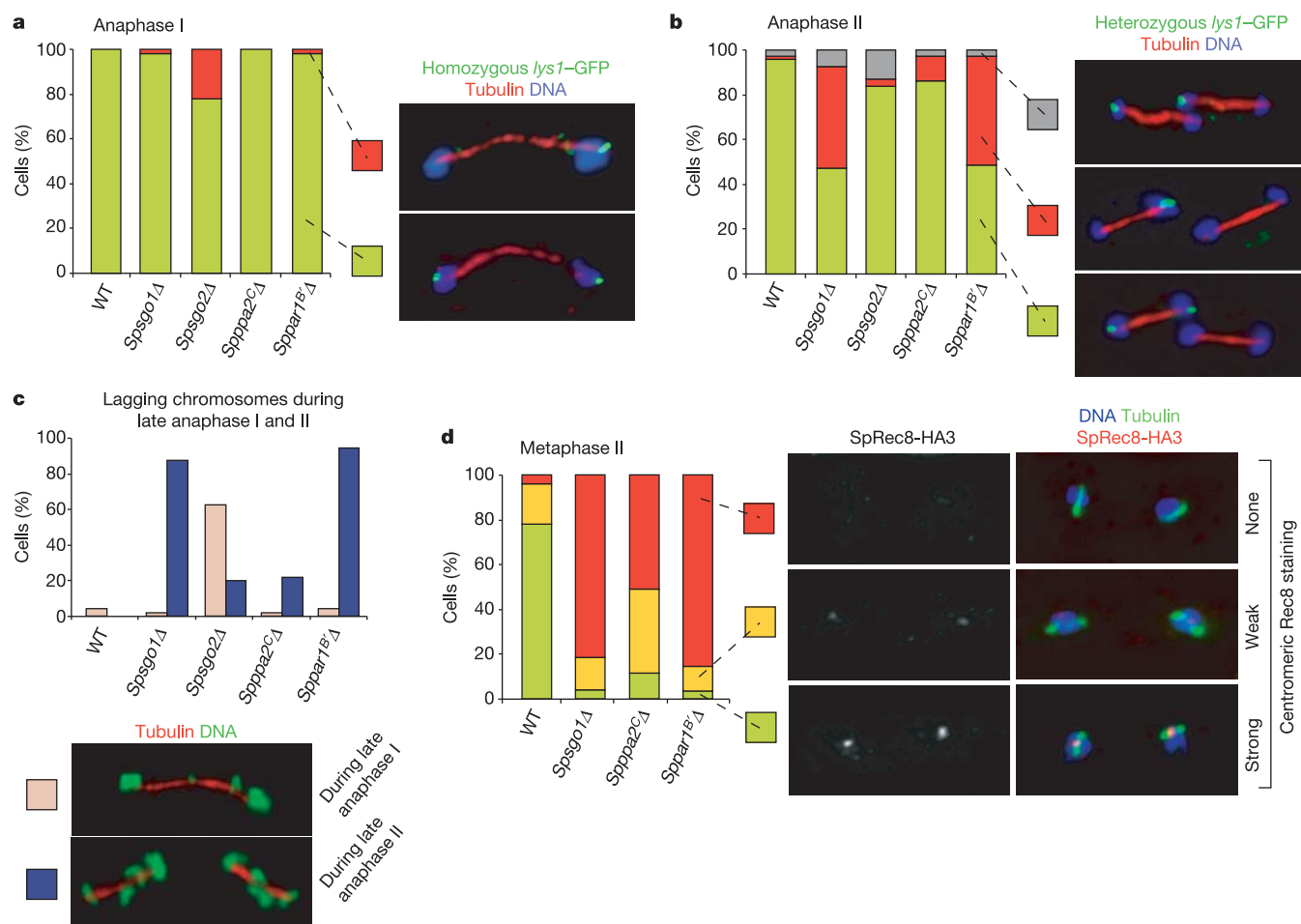


Figure 2 | PP2A- π is required to protect centromeric sister chromatid cohesion at meiosis I in *S. pombe*. **a**, **b**, Segregation of homozygous **(a)** or heterozygous **(b)** *lys1*-GFP in late anaphase I **(a)** or anaphase II **(b)** cells. *h⁹⁰* **(a)** or *h⁹⁰/h⁹⁰* diploid **(b)** wild-type, *Spsgo1 Δ* , *Spsgo2 Δ* , *Spppa2^C Δ* and *Sppar1^{B'} Δ* cells were sporulated for 18 h in liquid medium and samples were taken for *in situ* immunofluorescence microscopy ($n = 100$). Late anaphase I or anaphase II cells were identified as binucleates or tetranucleates containing one or two bipolar spindles of more than 7 or 4.5 μ m in length,

respectively. **c**, The same cells analysed in **b** were scored for the presence of lagging chromosomes in late anaphase I and anaphase II ($n = 100$). **d**, *h⁹⁰* wild-type, *Spsgo1 Δ* , *Spppa2^C Δ* and *Sppar1^{B'} Δ* cells, expressing *SpRec8*-HA3, were sporulated for 20 h in liquid medium and samples were taken for *in situ* immunofluorescence analysis of *SpRec8* at metaphase II centromeres ($n = 100$). Metaphase II cells were identified as binucleates containing two bipolar spindles.

Scub1Δ or *Scrts1^{B'}Δ* cells and very faint in the remaining 20% (Fig. 3d). Whole-cell staining of ScRec8 in *Scrts1^{B'}Δ* cells revealed a similar picture. Few if any *Scrts1^{B'}Δ* metaphase II binucleates contained centromeric ScRec8 (Supplementary Fig. S3d). Notably, deletion of *ScRTS1^{B'}* had no effect on the distribution of ScRec8 on chromosomes before meiosis I (Supplementary Fig. S3e, f).

Persistence of centromeric sister chromatid cohesion can be directly measured in *S. cerevisiae* using monopolin (for example, *Scmam1*) mutants, which attempt to pull sister centromeres to opposite poles at meiosis I but are prevented from disjoining them by cohesin that has resisted attack by separase³. This causes accumulation of uninucleate cells with bipolar spindles and low levels of securin (ScPds1). Deletion of *ScRTS1^{B'}* (but not *ScSPO11*) suppresses this meiosis I division defect (Fig. 3e) as does deletion of *ScSgo1* (ref. 10). However, owing to the presence of chiasmata,

which still enables co-orientation of some homologous centromeres even in the absence of monopolin, sister centromeres segregated to opposite poles in only 40% of cells (Fig. 3f). Elimination of crossing over by deleting *ScSPO11* enabled 85% of *Scrts1^{B'}Δ Scmam1Δ* cells to undergo a fully equational division. These results imply that PP2A- π actively maintains centromeric sister chromatid cohesion between meiotic divisions in *S. cerevisiae*.

Association of PP2A- π with centromeres depends on Sgo1

Unlike SpSgo1, most SpPP2A- π is not associated with chromosomes and is distributed throughout cells undergoing meiosis I (data not shown), which is consistent with our finding that most SpPP2A- π is not actually associated with SpSgo1 (Fig. 1a, b; see also Supplementary Table S1). This suggests that SpSgo1 mediates protection of centromeric sister chromatid cohesion by recruiting a small but

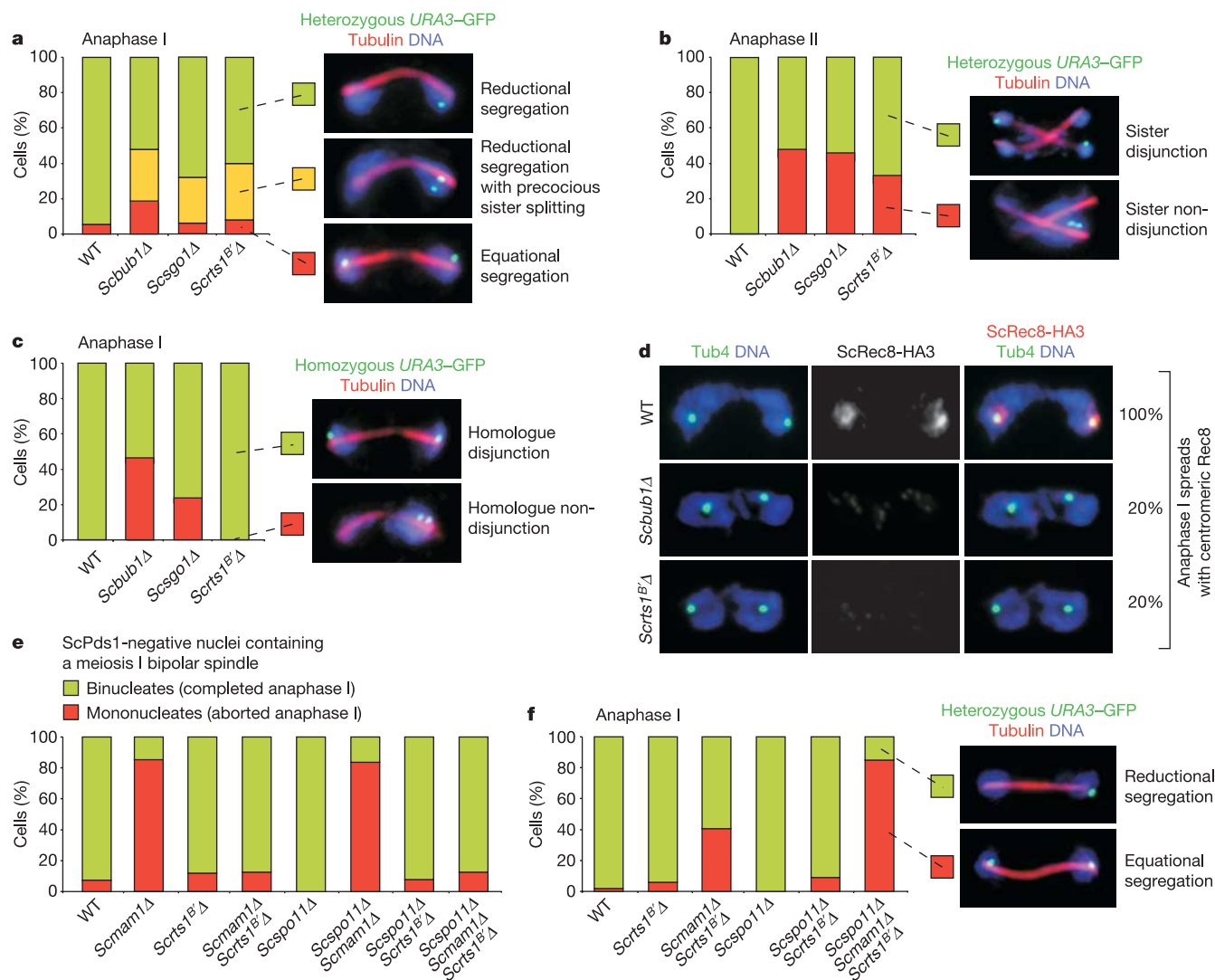


Figure 3 | PP2A- π prevents premature loss of centromeric sister chromatid cohesion at meiosis I in *S. cerevisiae*. **a, b**, Homozygous diploid wild-type, *Scub1Δ*, *Scsgo1Δ* and *Scrts1^{B'}Δ* cells, heterozygous for *URA3-GFP* and containing ScPds1-myc18, were sporulated in liquid culture for 6 h and samples taken for *in situ* immunofluorescence staining. The segregation of *URA3-GFP* at anaphase I (**a**) or anaphase II (**b**) was monitored in cells containing low nuclear levels of ScPds1-myc18 and either one (**a**) or two (**b**) bipolar spindles in binucleates or tetranucleates, respectively. **c**, As in **a**, except that wild-type, *Scub1Δ*, *Scsgo1Δ* and *Scrts1^{B'}Δ* cells were homozygous for *URA3-GFP*. **d**, Wild-type, *Scub1Δ* and *Scrts1^{B'}Δ* cells, taken from the same experiment as in **a**, were analysed for the presence of ScRec8-HA3 on chromosome spreads. Shown are typical examples of Rec8 staining in

anaphase I nuclei, in which a single spindle pole body (marked by ScTub4 staining) was observed in each lobe of a bilobed chromatin mass. The percentage of such nuclei containing ScRec8-HA3 co-localizing with the spindle pole bodies (centromeric Rec8) was scored ($n = 15$).

e, f, Homozygous diploid wild-type and mutant strains, carrying ScPds1-myc18 and heterozygous for *URA3-GFP*, were sporulated in liquid medium and samples taken for *in situ* immunofluorescence staining 6 h after resuspension of cells in sporulation medium. The ability of cells to undergo anaphase I was monitored in such cells containing low nuclear ScPds1-myc18 levels and a single bipolar spindle (**e**) ($n = 100$). In such strains that can efficiently undergo anaphase I, the segregation pattern of *URA3-GFP* was monitored (**f**) ($n = 100$).

vitaly important fraction of PP2A- π to meiosis I centromeres and not *vice versa*. As predicted by this hypothesis, localization of SpSgo1 to metaphase I centromeres was unaffected by *Sppar1^{B'}* Δ or *Spppa2^C* Δ (Fig. 4a), although it was abolished by *Spub1 Δ* (ref. 8). To

detect chromosomal SpPP2A- π , cells undergoing meiosis I were fixed with formaldehyde, and sheared DNA immunoprecipitated with epitope tags on SpPar1^{B'}, SpPpa2^C, SpSgo1 or SpRec8 was hybridized to a high-density oligonucleotide array covering chromosomes 2 and 3

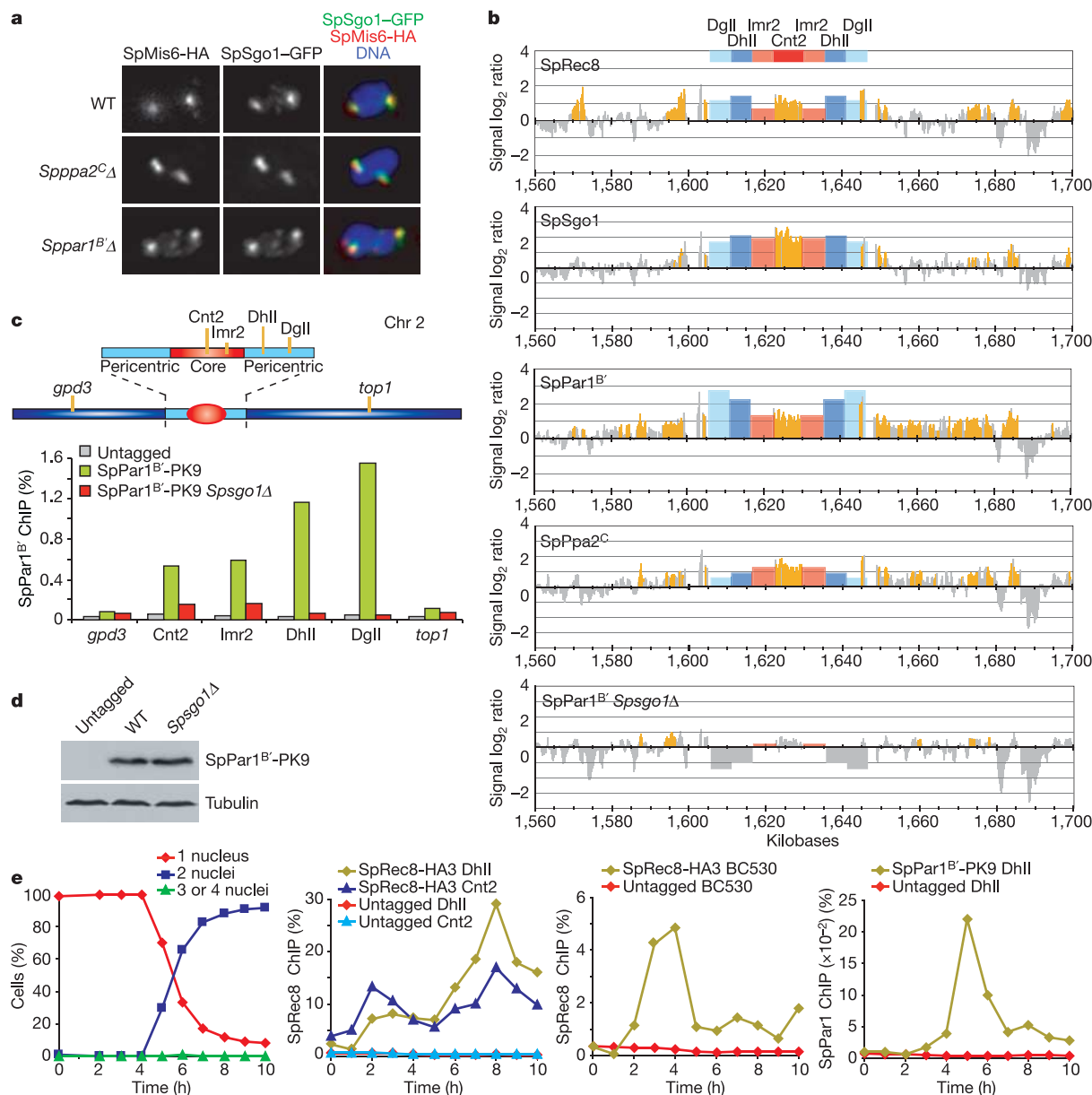


Figure 4 | PP2A- π localizes to *S. pombe* centromeres in an SpSgo1-dependent manner. **a**, SpSgo1 localizes to centromeres independently of PP2A- π . *h⁹⁰* wild-type, *Spppa2^C* Δ and *Sppar1^{B'}* Δ cells producing SpSgo1-GFP and the kinetochore protein SpMis6-HA3 were sporulated for 16 h in PMG-N media and samples taken. Co-localization of SpSgo1-GFP and SpMis6-HA3 was analysed by *in situ* immunofluorescence. **b**, *Sppat1-114* homozygous diploid cells carrying SpRec8-HA3, SpSgo1-PK9, SpPar1^{B'}-PK9, SpPpa2^C-HA3 and SpPar1^{B'}-PK9 *Spsgo1* Δ were harvested 4 h after induction of synchronous meiosis (that is, before anaphase I). The distribution of epitope-tagged proteins on chromosome 2 and most of chromosome 3 was analysed by ChIP and hybridization on a high-density oligonucleotide microarray. Only a 140-kb region surrounding the centromere of chromosome 2 (1,560–1,700 kb from the left telomere) is shown. The orange-shaded areas are regions containing significant enrichment of immunoprecipitated material, as detailed previously²⁷. Grey-shaded areas represent signals that are not statistically significant. Signals for the DgII, DhII and Imr2 repetitive centromere regions (light blue, dark blue and light red boxes, respectively) were determined by ChIP experiments

using quantitative real-time PCR (qRT-PCR) and calibrated to match the profile. For the complete microarray profiles showing entire chromosome 2 and most of chromosome 3 see Supplementary Figs S4 and S5. **c**, *Sppat1-114* homozygous diploid wild-type, SpPar1^{B'}-PK9 and SpPar1^{B'}-PK9 *Spsgo1* Δ cells were harvested 4 h after induction of synchronous meiosis. ChIP of various loci of chromosome 2, using qRT-PCR, was performed. The schematic drawing in the top panel illustrates the positions of amplified sequences. Both *gpd3* and *top1* are arm loci. **d**, Samples taken from the same cells as shown in **c** were analysed by western blotting for the expression levels of SpPar1^{B'}. Tubulin detection served as a loading control. **e**, *Spmes1-B44 Sppat1-114* homozygous diploid cells, producing the epitope-tagged proteins SpPar1^{B'}-PK9 and SpRec8-HA3, were induced to undergo meiosis I synchronously, eventually arresting before meiosis II. Chromatin binding of SpPar1^{B'}-PK9 and SpRec8-HA3 at various time points was analysed by ChIP and qRT-PCR. Amplified regions included the inner (Cnt2) and outer (DhII) centromeres, as well as a cohesion-associated chromosomal arm region (BC530).

(Fig. 4b; see also Supplementary Fig. S4). Association of proteins with the repetitive centromere regions, which are difficult to quantify by this method, was measured using a real-time polymerase chain reaction (PCR) assay. SpRec8 localized to inner (Cnt2 and Imr2) and outer (DhII and DgII) centromere regions, as well as to regions of convergent transcription along chromosome arms (Fig. 4b; see also Supplementary Figs S4 and S6). SpSgo1 localized to both inner and outer centromere regions, as did both SpPar1^{B'} and SpPpa2^C (Fig. 4b; see also Supplementary Figs S4 and S6). SpPar1^{B'} and SpPpa2^C might be more broadly distributed around centromeres than SpSgo1 (Fig. 4b).

To address whether SpPP2A- π persists at centromeres after meiosis I has been completed, we measured association of SpPar1^{B'}

and SpRec8 with chromosomes as *mes1-B44* mutants undergo meiosis I and arrest before the onset of meiosis II^{24,25}. Whereas SpRec8 was absent from arm sequences after meiosis I was completed, it persisted or even increased at centromeres (Fig. 4e). SpPar1^{B'} associated with centromeres shortly before the onset of meiosis I but disappeared when cells arrested at the onset of meiosis II. Crucially, the amount of SpPar1^{B'} associated with inner and outer centromere regions (but not the level of total protein; Fig. 4d) was severely reduced in *Spsgo1* Δ cells during meiosis I (Fig. 4b, c). Chromatin immunoprecipitations (ChIPs) hybridized to microarrays suggested that centromeric SpPar1^{B'} did not re-distribute to other chromosomal regions in *Spsgo1* Δ cells (Supplementary Fig. S5).

The distributions of PP2A- π , ScSgo1, ScBub1 and ScRec8 on

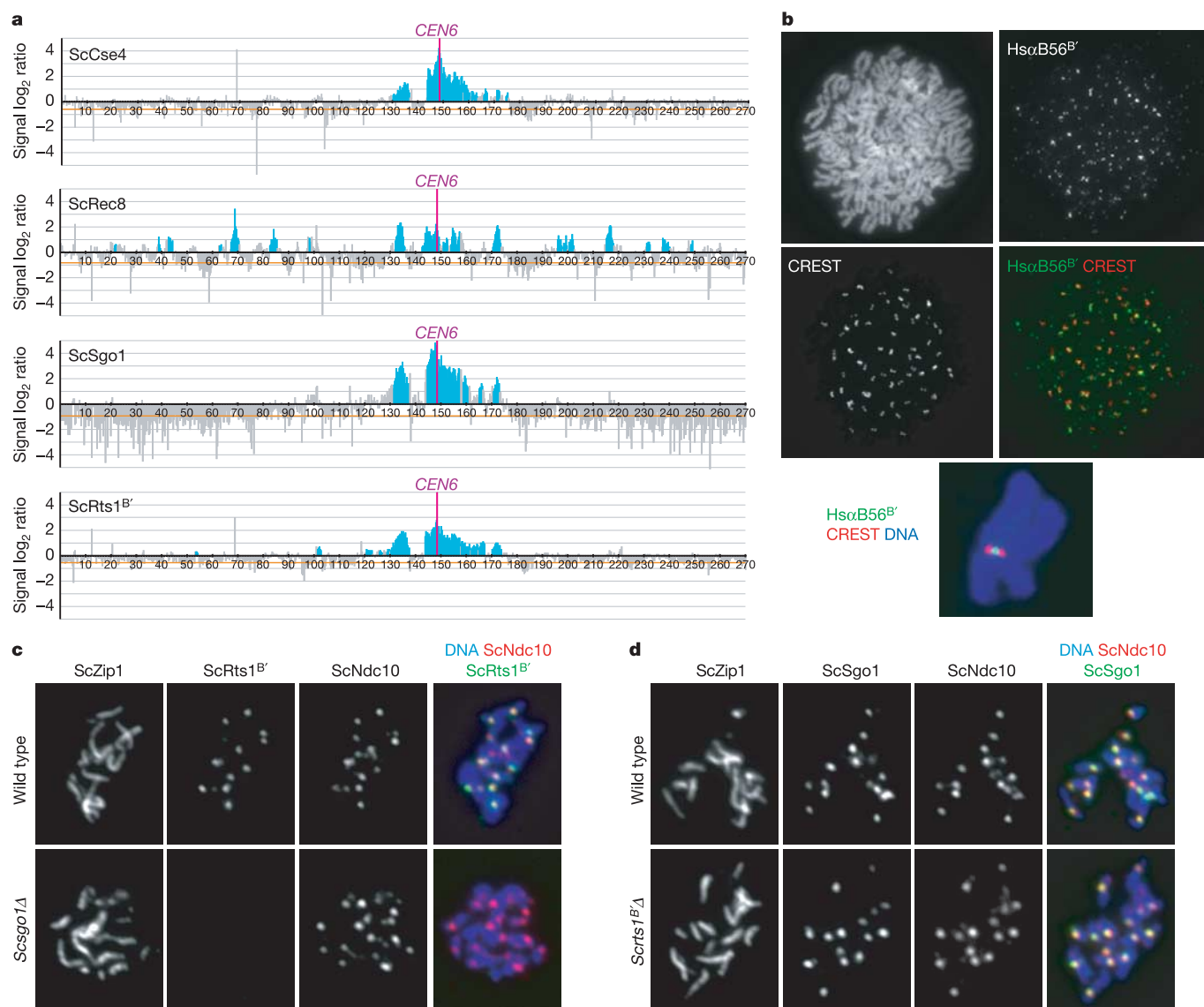


Figure 5 | PP2A- π is recruited to the centromeres in *S. cerevisiae* and mammalian cells. **a, Meiotic cells producing ScCse4-HA6, ScRec8-HA3, ScSgo1-PK6 and ScRts1^{B'}-PK3 were harvested 4 h after shifting cells to sporulation medium (that is, before the first meiotic division). Benomyl (30 $\mu\text{g ml}^{-1}$) was added 1 h after transfer to ensure that cells did not progress past meiosis I. The distribution of the tagged proteins on chromosome VI was analysed by ChIP and hybridization on a high-density oligonucleotide microarray. The blue-shaded areas are regions of the chromosome that contain significant enrichment of immunoprecipitated material, as detailed previously²⁷. Grey-shaded areas represent signals that are not statistically significant. The orange horizontal line indicates the average signal ratio of**

loci not enriched in the immunoprecipitated fraction. **b**, Distribution of Hs α B56^{B'} on HeLa chromosomes. Shown is a typical chromosome spread of a mitotic cell immunostained to detect endogenous Hs α B56^{B'} and the centromere marker CREST. The arrow points to a single chromosome pair, of which an enlargement is shown in the lowest panel. **c**, **d**, Meiotic diploid wild-type, *Scsgo1* Δ and *ScRts1*^B Δ cells were harvested 5 h after transfer to sporulation medium and the localization of ScRts1^{B'}-myc9 (**c**) and ScSgo1-myc9 (**d**) on late pachytene chromosome spreads was analysed. All spreads were co-stained for ScNdc10-HA3 (used as a marker for centromeres) and the synaptonemal complex protein ScZip1.

meiotic *S. cerevisiae* chromosomes were analysed by hybridizing chromatin immunoprecipitates to chromosome VI oligonucleotide microarrays^{26,27} (Fig. 5a; see also Supplementary Fig. S7a). The ScSgo1, ScBub1, ScRts1^{B'} and ScPph22^C distributions were similar to those of two kinetochore proteins, CENP-A (ScCse4) and ScNdc10. All six proteins localized to a large (~45-kilobase (kb)) region surrounding the 125-base-pair *CEN6* sequence that confers centromere function. This result was surprising because previous studies concluded that ScCse4 and ScNdc10 are restricted to a single centromeric nucleosome^{28,29}. The extended 'centromeric' region marked by these kinetochore proteins contained several prominent and closely spaced peaks of ScRec8, which was also detected at regions of convergent

transcription along chromosome arms (Fig. 5a; see also refs 27, 30). We suggest that the fraction of ScRec8 that co-localizes with ScPP2A- π within the 45-kb centromere region is resistant to separase cleavage at the onset of anaphase I, whereas the fraction of ScRec8 along chromosome arms that does not co-localize with ScPP2A- π is susceptible. ScCse4 and ScNdc10, as well as ScPP2A- π , ScSgo1 and ScBub1, were found in a narrower window (~20 kb) around *CEN6* in mitotic cells (Supplementary Fig. S7b).

We also observed ScPP2A- π at centromeres on chromosome spreads of late pachytene cells (Fig. 5c). The concentration of ScRts1^{B'} at centromeres depended on ScSgo1 and ScBub1, whereas that of ScSgo1 depended on ScBub1 but not ScRts1^{B'} (Fig. 5c, d; see

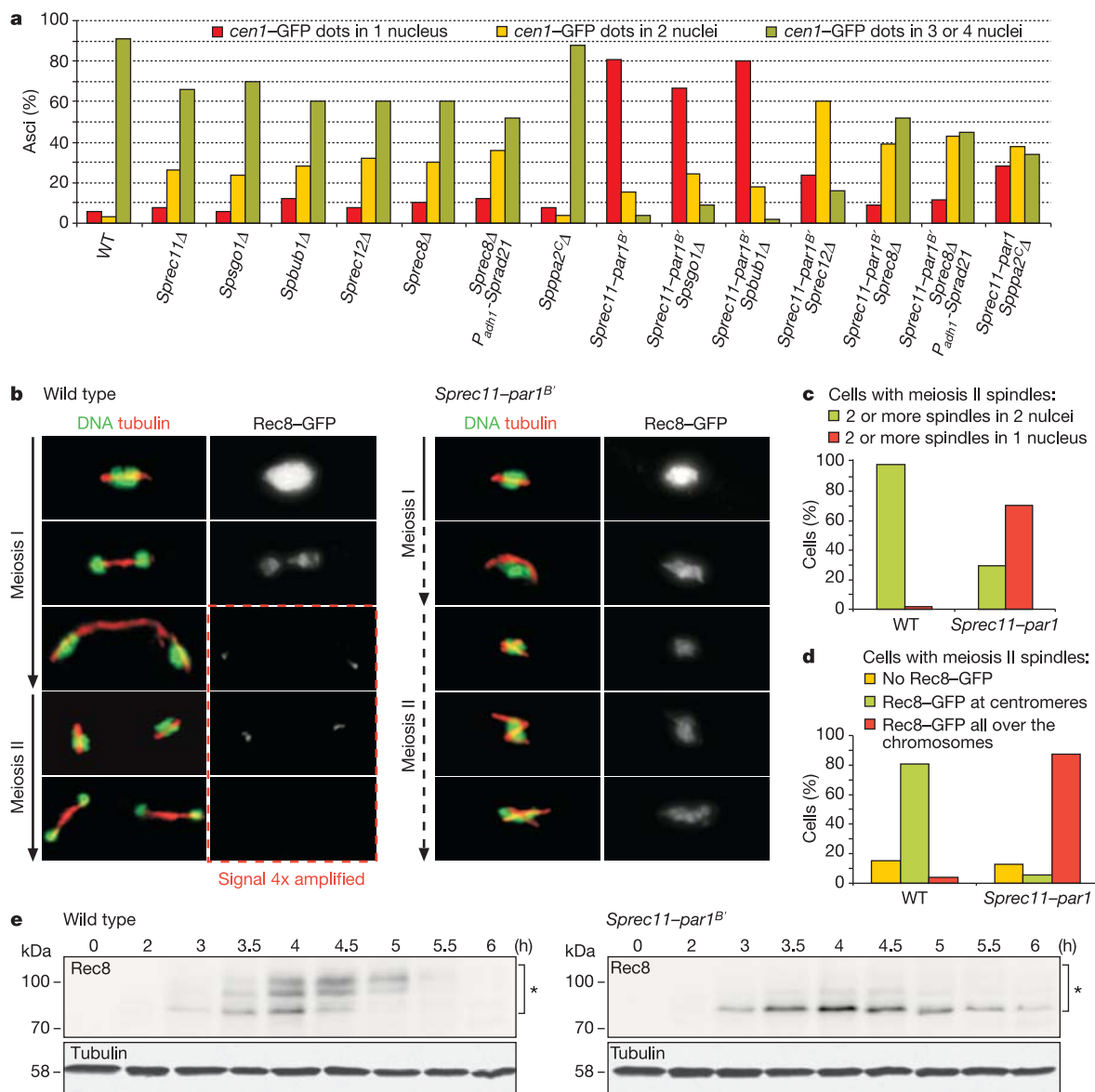


Figure 6 | Artificial recruitment of PP2A- π to chromosome arms in *S. pombe* prevents removal of arm cohesin, blocks nuclear divisions and causes dephosphorylation of SpRec8 during meiosis. **a, *h⁹⁰* cells, expressing the SpRec11-HA3-Par1^{B'} fusion protein from the endogenous *SpRec11* locus, together with control strains as shown, were sporulated for 40 h on PMG-N plates. Nuclear division was monitored by scoring the segregation of homozygous *lys1*-GFP into separate nuclei ($n = 100$). This was done to avoid scoring of nuclear fragmentation as divisions, as they frequently occur upon blocked chromosome segregation⁶. **b**, *h⁹⁰* cells, expressing either SpRec8-GFP or both SpRec8-GFP and the SpRec11-HA3-Par1^{B'} fusion protein, were sporulated for 20 h in PMG-N medium and samples were**

taken. Cells were stained for the presence of SpRec8-GFP and tubulin. Representative cells in various meiotic stages are shown. **c**, Cells from **b** that contained two or more spindles were scored for the presence of a divided nucleus ($n = 100$). **d**, Cells from **b** that contained two or more spindles were scored for the presence and localization of SpRec8-GFP ($n = 100$). **e**, *SpPat1-114* homozygous diploid cells producing either SpCut2-myc13 or both SpCut2-myc13 and the SpRec11-HA3-Par1^{B'} fusion protein were sporulated synchronously with samples taken at the indicated time points after induction of meiosis. Samples were analysed by western blotting for the presence of SpRec8 and tubulin. The asterisk denotes the migration of SpRec8 including the various slower-migrating phosphorylated forms of the protein.

also Supplementary Fig. S8a, b). ScBub1's association with centromeres was independent of both ScSgo1 and ScRts1^{B'} (Supplementary Fig. S8c). These data suggest that ScBub1 recruits ScSgo1 to centromeres, which in turn recruits PP2A- π . Interestingly, both ScSgo1 and ScRts1^{B'} were found at low levels along the length of most chromosomes in *Scbub1 Δ* mutants and in 'kinase dead' *Scbub1(K733R)* mutants (Supplementary Fig. S8a, b). Thus, the kinase activity of ScBub1 seems to be required to restrict the chromatin association of both ScSgo1 and PP2A- π to centromeric regions.

The distribution of mammalian PP2A- π was also analysed on chromosome spreads. The endogenous α -isoform of the human B' subunit (Hs α B56^{B'}) co-localized with centromeres marked by CREST in HeLa cells (Fig. 5b). Its depletion by RNA interference (RNAi) greatly reduced the protein level of Hs α B56^{B'} and the number of Hs α B56^{B'} foci co-localizing with CREST (Supplementary Fig. S9a, b), showing that the antibody used for detecting Hs α B56^{B'} was specific to this isoform. In addition, overexpressed epitope-tagged versions of α -, β -, γ -, δ - and ϵ -isoforms of HsB56^{B'} were also observed to co-localize with CREST in chromosome spreads from stably transfected NIH3T3 cells (Supplementary Fig. S9c).

Is PP2A- π sufficient to confer protection of Rec8?

If the sole role of Bub1 and Sgo1, with regards to their function in protecting centromeric cohesion, is to recruit PP2A- π to centromeres, and the role of PP2A- π is to remove phosphate groups from proteins whose phosphorylation is required for Rec8's cleavage, then the artificial recruitment of PP2A- π to chromosome arms should inhibit Rec8 cleavage at this location and thereby hinder the first meiotic division. To test this, we fused SpPar1^{B'} to the C terminus of the meiosis-specific Scc3-like cohesin subunit SpRec11, which is essential for sister chromatid cohesion along chromosome arms³¹. Upon transferring *h*⁹⁰ cells to sporulation conditions, SpRec11-Par1^{B'} accumulated on chromosomes and co-localized with SpRec8 during meiosis I (Supplementary Fig. S10a). Most cells failed to undergo the first meiotic division but nevertheless produced asci containing either one or more unequally sized spores (Fig. 6a). Their lack of chromosome segregation was not caused by a lack of meiotic progression because cells formed first meiosis I and subsequently meiosis II bipolar spindles. Consequently, mononucleate cells with multiple spindles spanning four spindle pole bodies could be seen with high frequency (Fig. 6b, c). *Sprec11-par1^{B'}* cells with meiosis II spindles contained SpRec8 throughout the chromosomes (and not merely at centromeres, as occurs in wild type) (Fig. 6b, d). Elimination of crossing over by deleting *Sprec12* permitted most *Sprec11-par1^{B'}* cells to undergo one, but not a second, division (Fig. 6a), suggesting that their meiosis I defect may be in resolving chiasmata. Crucially, deleting *Sprec8* enabled *Sprec11-par1^{B'}* cells to undergo two divisions, even when they expressed high levels of SpRec8's counterpart SpRad21. Because the lack of meiotic divisions caused by SpRec11-Par1^{B'} was largely, but not completely, suppressed by deleting *Spppa2^C* (Fig. 6a), we suggest that the phenotype is caused by the abnormal recruitment to chromosome arms of SpPP2A- π 's catalytic activity. We propose that this blocks chiasmata resolution by blocking SpRec8 (but presumably not SpRad21) cleavage. Notably, SpPP2A- π recruited directly to chromosome arms in this fashion blocked meiosis I independently of both SpSgo1 and SpBub1 (Fig. 6a). This suggests that neither of these proteins is essential for the catalytic activity of PP2A- π in preventing Rec8 cleavage once PP2A- π has been recruited to chromosomes.

To investigate whether SpPP2A- π causes dephosphorylation of Rec8, we compared its electrophoretic mobility in *Sprec11* and *Sprec11-par1^{B'}* cells as they undergo synchronous meiosis triggered by SpPat1 inactivation. Under these conditions, *Sprec11-par1^{B'}* had a less severe but still significant effect on meiosis I (data not shown). It did not, however, affect pre-meiotic DNA replication or delay securin degradation (Supplementary Fig. S10b, c). Importantly,

fusion of SpRec11 to SpPar1^{B'} largely eliminated the formation of slow-migrating forms of SpRec8 (Fig. 6e). These forms normally appear shortly before the degradation of SpRec8 and have previously been shown to be caused by phosphorylation³².

Discussion

The preservation of centromeric sister chromatid cohesion during meiosis I is a crucial aspect of meiosis. Defects in this process might contribute to the precocious separation of sister chromatids observed in human oocytes^{33–35} that might in turn cause fetal aneuploidy and thereby Down's syndrome. Our work suggests that Bub1 and Sgo1 perform this task in both fission and budding yeast by recruiting to meiosis I centromeres a specific form of PP2A (PP2A- π). Because transient treatment of mouse oocytes with okadaic acid, a potent inhibitor of PP2A, also causes the precocious separation of sister centromeres³⁶, we suggest that this mechanism might be conserved between fungi and mammals. Bub1, Sgo1 and PP2A- π co-localize with kinetochore-specific nucleosomes in *S. cerevisiae* in a surprisingly large interval surrounding the core centromere-determining sequence *CEN6*, which suggests that the centromeres in this organism might be much larger than previously suspected.

We suggest that PP2A- π protects centromeric cohesin from separase activity by dephosphorylating proteins whose phosphorylation is required for efficient Rec8 cleavage, such as separase or Rec8 itself. The finding that neither ScScc1 nor SpRad21 can be protected from separase activity by Sgo1 proteins^{3,37} suggests that Rec8 might be the key target of PP2A- π . Consistent with this hypothesis is the observation that SpPP2A- π can prevent Rec8 removal from chromosomes and causes its dephosphorylation when artificially recruited to chromosome arms. Phosphorylation of Rec8, possibly by Polo-like kinases^{38–40}, might promote cleavage whereas dephosphorylation by PP2A- π might hinder it. Our finding that Sgo1 is associated with PP2A- π in mitotic HeLa cells raises the possibility that Sgo1 protects centromeric cohesin from the prophase pathway during mitosis by a similar mechanism. In this case, the target of PP2A- π is likely to be cohesin's Scc3/SA2 subunit^{13,14}, the phosphorylation of which causes cohesin to dissociate from chromosomes in the absence of α -kleisin cleavage. It is also possible that PP2A- π hinders removal of cohesin from meiotic centromeres by a prophase pathway⁴¹.

METHODS

Yeast genetics and molecular biology. Deletion or replacement of *S. pombe* and *S. cerevisiae* genes was performed as detailed in Supplementary Information. Genotypes of strains used are listed in Supplementary Table S2. Induction of meiosis in diploid *S. pombe pat1-114* cells was performed as described⁷. Sporulation of diploid *S. cerevisiae* in liquid culture was performed essentially as previously published⁵.

Mammalian cell lines and RNAi. Mouse Sgo1-TAP was expressed in HeLa cells from a bacterial artificial chromosome (RP24-185F17) integrated into the genome of HeLa cells²². Mouse NIH3T3 cell lines, stably expressing HA4-tagged human α -, β -, γ -, δ - and ϵ -isoforms of B56, were obtained by transfection⁴². RNAi was performed by standard methods using oligonucleotides 5'-GUUCUU AUUCCUAUGCAUA-3' or 5'-CAGCUUGCCUCUCAAUUCA-3'. See Supplementary Information for further details.

TAP and mass spectrometry. TAP-tagged proteins were purified from yeast and mammalian cultures essentially as previously described⁴³, with minor modifications. Protein samples were trypsin digested, and the resulting peptides separated via nano-capillary liquid chromatography and identified by online tandem mass spectrometry. See Supplementary Information for details.

Immunostaining of whole cells and chromosome spreads. *S. pombe* and *S. cerevisiae* cells were fixed and stained for immunofluorescence microscopy essentially as described^{7,10}. *S. cerevisiae*, HeLa and NIH3T3 chromosome spreads were performed and immunostained as described previously^{10,44}.

Chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) and hybridization to Affymetrix high-density oligonucleotide arrays of *S. cerevisiae* chromosome VI and *S. pombe* chromosomes 2 and 3 were performed essentially as previously described^{26,27}. ChIP on repetitive *S. pombe* centromere sequences by quantitative real-time PCR was performed essentially as described previously⁴⁵. See Supplementary Information for details.

Received 13 October 2005; accepted 20 February 2006.

Published online 15 March 2006.

1. Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* **74**, 595–648 (2005).
2. Nasmyth, K. How might cohesin hold sister chromatids together? *Phil. Trans. R. Soc. Lond. B* **360**, 483–496 (2005).
3. Toth, A. *et al.* Functional genomics identifies monopolin: a kinetochore protein required for segregation of homologs during meiosis I. *Cell* **103**, 1155–1168 (2000).
4. Rabitsch, K. P. *et al.* Kinetochore recruitment of two nucleolar proteins is required for homolog segregation in meiosis I. *Dev. Cell* **4**, 535–548 (2003).
5. Buonomo, S. B. *et al.* Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by separin. *Cell* **103**, 387–398 (2000).
6. Kitajima, T. S., Miyazaki, Y., Yamamoto, M. & Watanabe, Y. Rec8 cleavage by separase is required for meiotic nuclear divisions in fission yeast. *EMBO J.* **22**, 5643–5653 (2003).
7. Rabitsch, K. P. *et al.* Two fission yeast homologs of *Drosophila* Mei-S332 are required for chromosome segregation during meiosis I and II. *Curr. Biol.* **14**, 287–301 (2004).
8. Kitajima, T. S., Kawashima, S. A. & Watanabe, Y. The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. *Nature* **427**, 510–517 (2004).
9. Kerrebrock, A. W., Moore, D. P., Wu, J. S. & Orr-Weaver, T. L. Mei-S332, a *Drosophila* protein required for sister-chromatid cohesion, can localize to meiotic centromere regions. *Cell* **83**, 247–256 (1995).
10. Katis, V. L., Galova, M., Rabitsch, K. P., Gregan, J. & Nasmyth, K. Maintenance of cohesin at centromeres after meiosis I in budding yeast requires a kinetochore-associated protein related to MEI-S332. *Curr. Biol.* **14**, 560–572 (2004).
11. Marston, A. L., Tham, W. H., Shah, H. & Amon, A. A genome-wide screen identifies genes required for centromeric cohesion. *Science* **303**, 1367–1370 (2004).
12. Bernard, P., Maure, J. F. & Javerzat, J. P. Fission yeast Bub1 is essential in setting up the meiotic pattern of chromosome segregation. *Nature Cell Biol.* **3**, 522–526 (2001).
13. Hauf, S. *et al.* Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol.* **3**, e69 (2005).
14. McGuinness, B. E., Hirota, T., Kudo, N. R., Peters, J. M. & Nasmyth, K. Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. *PLoS Biol.* **3**, e86 (2005).
15. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
16. Kinoshita, N., Yamano, H., Niwa, H., Yoshida, T. & Yanagida, M. Negative regulation of mitosis by the fission yeast protein phosphatase ppa2. *Genes Dev.* **7**, 1059–1071 (1993).
17. Kinoshita, N., Ohkura, H. & Yanagida, M. Distinct, essential roles of type 1 and 2A protein phosphatases in the control of the fission yeast cell division cycle. *Cell* **63**, 405–415 (1990).
18. Kinoshita, K. *et al.* The regulatory subunits of fission yeast protein phosphatase 2A (PP2A) affect cell morphogenesis, cell wall synthesis and cytokinesis. *Genes Cells* **1**, 29–45 (1996).
19. Jiang, W. & Hallberg, R. L. Isolation and characterization of *par1⁺* and *par2⁺*: two *Schizosaccharomyces pombe* genes encoding B' subunits of protein phosphatase 2A. *Genetics* **154**, 1025–1038 (2000).
20. Ronne, H., Carlberg, M., Hu, G. Z. & Nehlin, J. O. Protein phosphatase 2A in *Saccharomyces cerevisiae*: effects on cell growth and bud morphogenesis. *Mol. Cell. Biol.* **11**, 4876–4884 (1991).
21. Zhao, Y., Boguslawski, G., Zitomer, R. S. & DePaoli-Roach, A. A. *Saccharomyces cerevisiae* homologs of mammalian B and B' subunits of protein phosphatase 2A direct the enzyme to distinct cellular functions. *J. Biol. Chem.* **272**, 8256–8262 (1997).
22. Kittler, R. *et al.* RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proc. Natl Acad. Sci. USA* **102**, 2396–2401 (2005).
23. Lew, D. J. & Burke, D. J. The spindle assembly and spindle position checkpoints. *Annu. Rev. Genet.* **37**, 251–282 (2003).
24. Izawa, D., Goto, M., Yamashita, A., Yamano, H. & Yamamoto, M. Fission yeast Mes1p ensures the onset of meiosis II by blocking degradation of cyclin Cdc13p. *Nature* **434**, 529–533 (2005).
25. Shimoda, C., Hirata, A., Kishida, M., Hashida, T. & Tanaka, K. Characterization of meiosis-deficient mutants by electron microscopy and mapping of four essential genes in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **200**, 252–257 (1985).
26. Katou, Y. *et al.* S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424**, 1078–1083 (2003).
27. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–578 (2004).
28. Meluh, P. B. & Koshland, D. Budding yeast centromere composition and assembly as revealed by *in vivo* cross-linking. *Genes Dev.* **11**, 3401–3412 (1997).
29. Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell* **94**, 607–613 (1998).
30. Glynn, E. F. *et al.* Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. *PLoS Biol.* **2**, e259 (2004).
31. Kitajima, T. S., Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Distinct cohesin complexes organize meiotic chromosome domains. *Science* **300**, 1152–1155 (2003).
32. Watanabe, Y. & Nurse, P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature* **400**, 461–464 (1999).
33. Angell, R. First-meiotic-division nondisjunction in human oocytes. *Am. J. Hum. Genet.* **61**, 23–32 (1997).
34. Kuliev, A., Cieslak, J., Ilkevitch, Y. & Verlinsky, Y. Chromosomal abnormalities in a series of 6,733 human oocytes in preimplantation diagnosis for age-related aneuploidies. *Reprod. Biomed. Online* **6**, 54–59 (2003).
35. Pellestor, F., Andreo, B., Arnal, F., Humeau, C. & Demaille, J. Maternal aging and chromosomal abnormalities: new data drawn from *in vitro* unfertilized human oocytes. *Hum. Genet.* **112**, 195–203 (2003).
36. Mailhes, J. B., Hilliard, C., Fuseler, J. W. & London, S. N. Okadaic acid, an inhibitor of protein phosphatase 1 and 2A, induces premature separation of sister chromatids during meiosis I and aneuploidy in mouse oocytes *in vitro*. *Chromosome Res.* **11**, 619–631 (2003).
37. Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Cohesins determine the attachment manner of kinetochores to spindle microtubules at meiosis I in fission yeast. *Mol. Cell. Biol.* **23**, 3965–3973 (2003).
38. Alexandru, G., Uhlmann, F., Mechtler, K., Poupart, M. A. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459–472 (2001).
39. Clyne, R. K. *et al.* Polo-like kinase Cdc5 promotes chiasmata formation and cosegregation of sister centromeres at meiosis I. *Nature Cell Biol.* **5**, 480–485 (2003).
40. Lee, B. H. & Amon, A. Role of Polo-like kinase *CDC5* in programming meiosis I chromosome segregation. *Science* **300**, 482–486 (2003).
41. Yu, H. G. & Koshland, D. Chromosome morphogenesis: condensin-dependent cohesin removal during meiosis. *Cell* **123**, 397–407 (2005).
42. McCright, B., Rivers, A. M., Audlin, S. & Virshup, D. M. The B56 family of protein phosphatase 2A (PP2A) regulatory subunits encodes differentiation-induced phosphoproteins that target PP2A to both nucleus and cytoplasm. *J. Biol. Chem.* **271**, 22081–22089 (1996).
43. Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
44. Peters, A. H., Plug, A. W., van Vugt, M. J. & de Boer, P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res.* **5**, 66–68 (1997).
45. Pidoux, A., Mellone, B. & Allshire, R. Analysis of chromatin in fission yeast. *Methods* **33**, 252–259 (2004).
46. Yamamoto, M. The molecular control mechanisms of meiosis in fission yeast. *Trends Biochem. Sci.* **21**, 18–22 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank K. Rabitsch, N. Kudo, C. Kraft, I. Steinmacher, R. Imre, K. Tanaka, R. Kittler, F. Stewart and A. Lorenz for help with experiments, as well as M. Yanagida, I. Poser, I. Hagan, K. Gull, K. Gould, C. Warren, F. Spencer, D. Virshup, W. Zachariae and Y. Watanabe for cell lines, strains and reagents. In addition we would like to thank members of the Nasmyth laboratory for discussions. J.G. was a recipient of the EMBO long-term fellowship. K.M. received funding from GenAU. E.O. was supported by a grant from the Austrian Science Foundation.

Author Contributions Experiments were designed and data analysed and interpreted by C.R., V.K. and K.N. *S. pombe* strain construction, cytological and biochemical analysis, TAP purification, conventional ChIP experiments and *in vitro* binding assays were performed by C.R. *S. cerevisiae* strain construction and cytological analysis was performed by V.K. ChIP using *S. pombe* and *S. cerevisiae* oligonucleotide microarrays was performed by Y.K., S.M., T.I. and K.S. *S. cerevisiae* TAP purification was performed by M.P. HeLa TAP purification was performed by J.G. from a cell line with help from F.B. and L.P. Cytological analysis of HeLa and NIH3T3 cells was performed by B.C., using cell lines generated by I.M. and E.O. Generation and characterization of human α B56 antibody was performed by I.M. and E.O. Mass spectrometry was performed by K.M. The manuscript was written by V.K., C.R. and K.N. W.H. and M.G. provided technical assistance.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.N. (kim.nasmyth@bioch.ox.ac.uk).

LETTERS

A regular period for Saturn's magnetic field that may track its internal rotation

G. Giampieri¹, M. K. Dougherty², E. J. Smith¹ & C. T. Russell³

The rotation rate of a planet is one of its fundamental properties. Saturn's rotation, however, is difficult to determine because there is no solid surface from which to time it, and the alternative 'clock'—the magnetic field—is nearly symmetrically aligned with the rotation axis^{1–7}. Radio emissions, thought to provide a proxy measure of the rotation of the magnetic field, have yielded estimates of the rotation period between 10 h 39 min 22 s and 10 h 45 min 45 s (refs 8–10). Because the period determined from radio measurements exhibits large time variations, even on time-scales of months, it has been uncertain whether the radio-emission periodicity coincides with the inner rotation rate of the planet. Here we report magnetic field measurements that revealed a time-stationary magnetic signal with a period of 10 h 47 min 6 s $\pm$ 40 s. The signal appears to be stable in period, amplitude and phase over 14 months of observations, pointing to a close connection with the conductive region inside the planet, although its interpretation as the 'true' inner rotation period is still uncertain.

A fundamental question concerning Saturn's interior remains unanswered despite measurements spanning a quarter of a century: what is the rotation rate of the deep interior? This is different from the case of Jupiter, where the internal rotation rate can be accurately measured thanks to the presence of a significant dipole tilt¹¹. In Saturn's case, where the magnetic field is highly axially symmetric, we have a dilemma: we need a clear azimuthal field component in order to measure the rotation rate of the planet, but in order to be able to invert the data to find the azimuthal field we need to know the rotation rate quite accurately, or else the azimuthal component will be averaged to zero. Periodic modulations in the magnetic field components that may be associated with the rotation of Saturn's interior were first found in the Pioneer 11 data, although a careful analysis of the 10.5-h periodicity observed seemed to rule out an internal origin¹². These periodic azimuthal signals continue to be intriguing because they hold promise of providing the necessary 'clock' to time Saturn's rotational period.

The lack of a significant azimuthal field of internal origin in Saturn's magnetic field has yet to be explained, and several planetary models have been proposed to account for it. The most notable is from Stevenson¹³, who proposed that the presence of a conductive shell, external to the dynamo region and differentially rotating with respect to it, can greatly suppress the non-axisymmetric components of the internal field away from the planet. In addition, data from the Pioneer and Voyager fly-bys were reanalysed using modern inversion techniques and without any *a priori* assumption about the symmetry of the internal field¹⁴. This analysis revealed the possible existence of some non-axisymmetric field, although the limited spatial sampling of the three fly-bys did not allow an unambiguous measurement of the higher-order components of the field to be made.

Recently, Cassini orbiter observations have greatly extended the *in situ* set of magnetic field observations at Saturn⁷. We have analysed

magnetic field data collected by the fluxgate magnetometer on board Cassini between Saturn orbit insertion (SOI, July 2004) and orbit 13 (August 2005), during which time the spacecraft flew by the planet on 15 separate occasions. The closest approach distance for these passes varies between 1.33 R_S and 6.19 R_S ($1R_S = 60,268$ km is defined as Saturn's planetary radius)¹⁵, and the data used were ten-minute averages inside 10 R_S . Best-fit parameters of an axially symmetric planetary field (up to octupole terms) and of an axially symmetric ring current field^{16,17} were derived, and a search for periodic modulations in the resulting residuals was carried out. The residuals of the three magnetic field components in a spherical polar co-ordinate system (B_r , B_θ , B_ϕ) can be seen in Fig. 1a–c respectively. The residuals reveal a clear periodic signature in all three of the components but most clearly in the azimuthal one (Fig. 1c), which is not affected at all

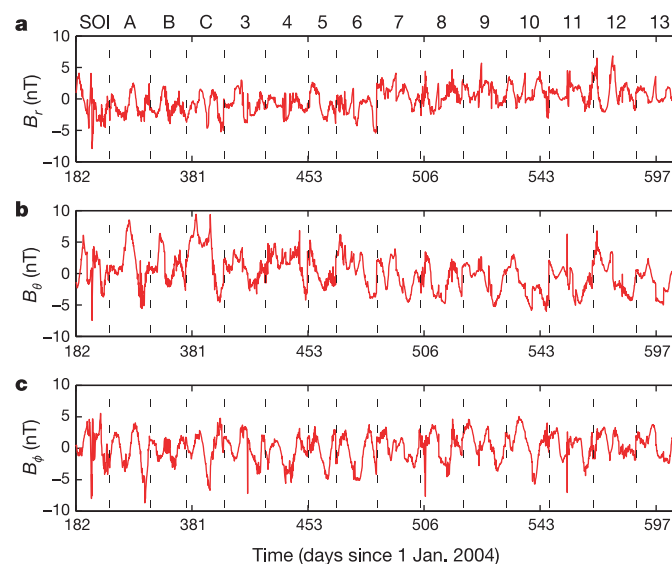


Figure 1 | Periodicity in Saturn's magnetic field. a–c, The spherical components of the magnetic field residuals (a, B_r ; b, B_θ ; c, B_ϕ) in the IAU reference frame. The vertical dashed lines separate the different passes, indicated in the top panel (from SOI to orbit 13). These residuals are calculated by subtracting the estimated contribution of the internal planetary field and the external magnetodisk current sheet. For each pass, best fit parameters of an axially symmetric magnetodisk field were derived. The magnetodisk parameters were assumed to respond to the time varying magnetospheric conditions, and therefore allowed to vary from pass to pass. In addition, the internal planetary field was estimated using standard inversion techniques on the entire sequence of observations. In order not to interfere with our search for periodic signatures in the data, the planetary field was assumed to be axisymmetric.

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. ²The Blackett Laboratory, Imperial College London, London SW7 2AZ, UK. ³UCLA Institute of Geophysics and Planetary Physics, Los Angeles, California 90025, USA.

by the removal of axisymmetric contributions from the magnetodisk and the internal field. This signal can allow the first accurate and direct measurement of the magnetic field periodicity.

A difficulty we face when searching for periodicities in time series data such as that in Fig. 1 is that the spacecraft is constantly moving with respect to the planet. This introduces a Doppler-like effect on the magnetic field, which is rotating with the planet; in particular, because Cassini is moving in the same sense as the planet rotates, the field is shifted towards longer periods. This effect, which is accumulated over many months and orbits of the planet, could make it impossible to determine the actual frequency of the periodic signal, whatever its origin. In order to remove this effect, we replace the actual physical time (t) of the observation with a 'longitudinal' time, which we define as $\tau = -\phi P/(2\pi)$, where ϕ is the IAU longitude of the spacecraft, and P is the IAU rotation period of Saturn (defined as $P = 10\text{ h } 39\text{ min } 22.4\text{ s}$; see ref. 18 for the definition of Saturn's IAU cartographic coordinates). The minus sign in the definition assures that τ is monotonically increasing, apart from a very short interval around closest approach during SOI when Cassini's velocity was larger than Saturn's angular velocity. The implication is that when Cassini is in a fixed position in inertial space, $\phi = -2\pi t/P$ and $\tau = t$. In general though, because of the orbital motion of the spacecraft, $\tau \leq t$. In such a scenario, a magnetic feature corotating with the planet would appear as a sinusoidal signal, which is proportional to $\exp(i2\pi\tau/P)$, independent of the spacecraft motion.

In addition, because we limit our analysis to data within the inner and middle magnetosphere, our observational time sequence contains data gaps that often last for many weeks. This makes the use of standard fast Fourier transform techniques impractical, and instead we use the Lomb-Scargle algorithm, which facilitates the analysis of unevenly spaced data¹⁹. Applying this spectral technique to the residual data in Fig. 1 produces the power spectra shown in Fig. 2. A very clear peak can be seen in all three of the components at $P = 10\text{ h } 47\text{ min } 6\text{ s} \pm 40\text{ s}$. No other peaks of this significance were found outside the plotted range. A series of side lobes extends on either side of the main peak, associated with an amplitude modulation of the signal due to Cassini's orbital motion. We can estimate an uncertainty of $\sigma_P = 40\text{ s}$ for this period from the half-width of the

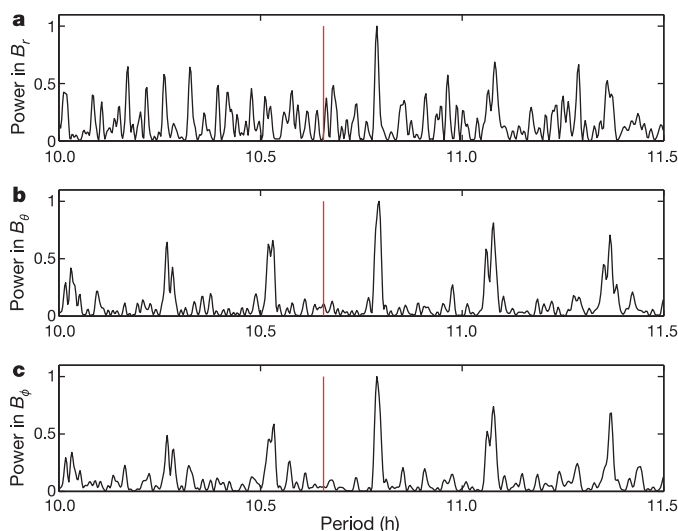


Figure 2 | Normalized spectra of the magnetic field residuals. a–c, The power spectra of the residuals from Fig. 1; the vertical red line indicates the IAU rotation period being used to date. No component at the IAU rotation period can be seen, whereas a peak at $10\text{ h } 47\text{ min } 6\text{ s}$ is clearly visible. Note the side lobes of the rotation peak, associated with the orbital period of Cassini. The width of the rotation peak, mainly due to the finite duration of the data record, limits the intrinsic time variation of the period to within 40 s.

peak in the spectrum. A simple inspection of the field components in the time domain reveals that the phase of the periodic signal varies by less than 0.5° from orbit to orbit.

Can this periodic signal be a direct signature of Saturn's rotation rate? If we accept that the radio emission period is related to the rotation rate of the planetary interior, then the implication is that the rotation of the planet has slowed down by several minutes over a 25-year period. Such a situation would require either a large external torque or a large redistribution in the internal density and velocity profiles, both unrealistic scenarios, because the dissipated energy ($\sim 10^{30}\text{ erg s}^{-1}$, under simple assumptions) would be a million times the planet's luminosity¹³. We must then conclude that the period measured in the radio emission is modulated by external effects. For example, the solar wind may cause the observed varying phase modulation of the radio source²⁰. On the other hand, the magnetic field signal, which we have only studied over the much shorter time span of the Cassini mission, shows no obvious variations in period, phase or amplitude, and can well be used to define an improved rotation rate of the planetary source. In this case, one should rescale the planetary longitude accordingly. In fact, only when the zero longitude axis is redefined by taking into account the new rotation period, are magnetic field observations, taken at the same longitude but at different epochs, found to overlap, as one would expect from an internal source.

The simplest model that produces a periodic field is the tilted dipole, but if this were the source of the periodic signal we would see associated changes that are not observed. More precisely, to generate a sinusoidal modulation proportional to $\cos\phi$ in the internal field, non-zero azimuthal terms of degree 1 are required in the magnetic field potential that depend on the other two spherical coordinates (r, θ), and no such latitudinal or radial dependence is observed in the residuals in Fig. 1. To see this more clearly, we show an example of the expected B_ϕ component, which would arise from a hypothetical tilted dipole (Fig. 3). This azimuthal field is very different from that which we observe. We conclude that a low-order non-axisymmetric

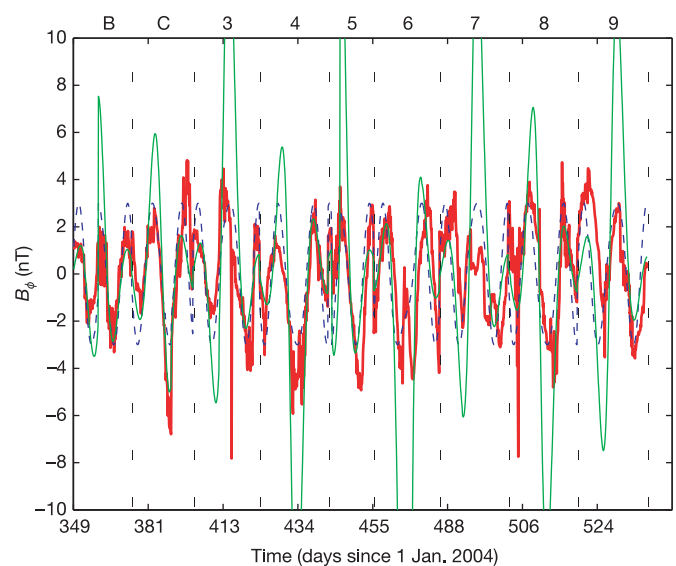


Figure 3 | Comparison with a tilted dipole signal. The expected azimuthal field arising from a tilted dipole is shown by the thin green curve, and is compared to the real data, in red, for orbit B through to orbit 9. Also shown is a sinusoid of period $10\text{ h } 47\text{ min } 6\text{ s}$ with a constant phase (blue dashed line). Note how the expected dipole signal is in phase with the azimuthal component of the magnetic field, thanks to the re-definition of the longitude from Cassini data. However, the increase in field values near closest approach cannot be found in the data itself. The assumed equatorial component of the dipole has a magnitude of 825 nT in the direction of -76° . As in Fig. 1, the vertical dashed lines separate the orbits.

internal field cannot be responsible for the large amplitude modulation shown in Fig. 1.

We note that local spin-periodic magnetic field perturbations observed in some of the Pioneer and Voyager data were explained via a 'camshaft' model, which proposed that a hypothetical anomaly (possibly magnetic or atmospheric) could generate periodic waves in the magnetosphere²¹. At present, we are looking at the polarization results from Cassini data in order to be able to confirm whether this hypothesis is able to explain the observed periodicity. We are left with the conundrum that we have a magnetic signal that appears to be stable in period, thus suggesting that it is linked to the rotation of the interior, but we do not yet know how that signal is produced.

Received 3 January; accepted 20 March 2006.

1. Smith, E. J. *et al.* Saturn's magnetic field and magnetosphere. *Science* **207**, 407–410 (1980).
2. Smith, E. J. *et al.* Saturn's magnetosphere and its interaction with the solar wind. *J. Geophys. Res.* **85**, 5655–5674 (1980).
3. Acuna, M. H. & Ness, N. F. The magnetic field of Saturn–Pioneer 11 observations. *Science* **207**, 444–446 (1980).
4. Acuna, M. H., Connerney, J. E. P. & Ness, N. F. Topology of Saturn's main magnetic field. *Nature* **292**, 721–724 (1981).
5. Ness, N. F. *et al.* Magnetic field studies by Voyager 1–Preliminary results at Saturn. *Science* **212**, 211–217 (1981).
6. Connerney, J. E. P., Ness, N. F. & Acuna, M. H. Zonal harmonic model of Saturn's magnetic field from Voyager 1 and 2 observations. *Nature* **298**, 44–46 (1982).
7. Dougherty, M. K. *et al.* Cassini magnetometer observations during Saturn Orbit Insertion. *Science* **307**, 1266–1270 (2005).
8. Desch, M. D. & Kaiser, M. L. Voyager measurement of the rotation period of Saturn's magnetic field. *Geophys. Res. Lett.* **8**, 253–256 (1981).
9. Gurnett, D. A. *et al.* Radio and plasma wave observations at Saturn from Cassini's approach and first orbit. *Science* **307**, 1255–1259 (2005).
10. Galopeau, P. H. M. & Lecacheux, A. Variations of Saturn's radio rotation period measured at kilometer wavelengths. *J. Geophys. Res.* **105**, 13089–13102 (2000).
11. Russell, C. T., Yu, Z. Y. & Kivelson, M. G. The rotation period of Jupiter. *Geophys. Res. Lett.* **28**, 1911–1912 (2001).
12. Espinosa, S. A. & Dougherty, M. K. Periodic perturbations in Saturn's magnetic field. *Geophys. Res. Lett.* **27**, 2785–2788 (2000).
13. Stevenson, D. J. Saturn's luminosity and magnetism. *Science* **208**, 746–748 (1980).
14. Giampieri, G. & Dougherty, M. K. Rotation rate of Saturn's interior from magnetic field observations. *Geophys. Res. Lett.* **31**, L16701, doi:10.1029/2004GL020194 (2004).
15. Cassini-Huygens-Operations-Saturn Tour Dates. (<http://saturn.jpl.nasa.gov/operations/cassini-calendar-ALL.cfm>) (accessed October 2005).
16. Connerney, J. E. P., Acuna, M. H. & Ness, N. F. Currents in Saturn's magnetosphere. *J. Geophys. Res.* **88**, 8779–8789 (1983).
17. Giampieri, G. & Dougherty, M. K. Modelling of the ring current in Saturn's magnetosphere. *Ann. Geophys.* **22**, 653–659 (2004).
18. Davies, M. E. *et al.* Report of the IAU/IAG/COSPAR working group on cartographic coordinates and rotational elements of the planets and satellites: 1994. *Celest. Mech. Dyn. Astron.* **63**, 127–148 (1996).
19. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ. Press, New York, 1992).
20. Cecconi, B. & Zarka, P. Model of a variable radio period for Saturn. *J. Geophys. Res.* **110**, A12203, doi:10.1029/2005JA011085 (2005).
21. Espinosa, S. A., Southwood, D. J. & Dougherty, M. K. How can Saturn impose its rotation period in a noncorotating magnetosphere? *J. Geophys. Res.* **108**, 1086, doi:10.1029/2001JA005084 (2003).

Acknowledgements Results presented here represent one aspect of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This research was performed while G.G. held a National Research Council senior research associateship award at JPL. We thank M. E. Burton and J. Wolf for suggestions on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.G. (giacomo.giampieri@jpl.nasa.gov).

Metallic transport in polyaniline

Kwanghee Lee^{1,2*}, Shinuk Cho¹, Sung Heum Park¹, A. J. Heeger², Chan-Woo Lee³ & Suck-Hyun Lee^{3*}

Despite nearly three decades of materials development, the transport properties in the ‘metallic state’ of the so-called conducting polymers are still not typical of conventional metals^{1–7}. The hallmark of metallic resistivity—a monotonic decrease in resistivity with temperature—has not been obtained at temperatures over the full range below room temperature; and a frequency dependent conductivity, $\sigma(\omega)$, typical of metals has also not been observed. In contrast, the low-temperature behaviour of ‘metallic’ polymers has, in all previous cases, exhibited an increase in resistivity as temperature is further decreased, as a result of disorder-induced localization of the charge carriers^{1–4}. This disorder-induced localization also changes the infrared response such that $\sigma(\omega)$ deviates from the prediction of Drude theory^{5–7}. Here we report classic metallic transport data obtained from truly metallic polymers. With polyaniline samples prepared using self-stabilized dispersion polymerization⁸, we find that for samples having room-temperature conductivities in excess of $1,000 \text{ S cm}^{-1}$, the resistivity decreases monotonically as the temperature is lowered down to 5 K, and that the infrared spectra are characteristic of the conventional Drude model even at the lowest frequencies measured.

The dynamics of the single-particle excitations in the partially filled conduction band of traditional metals are governed by coherent diffusion processes⁹. Electrons drift ballistically with a mean free path much longer than the structural repeat length and make random collisions with defects and with phonons, resulting in finite resistivity. As shown by Landau, the low-energy excitations in metals can be treated as non-interacting quasiparticles and described by Fermi-liquid theory¹⁰. This theory rationalizes and fundamentally explains the success of the free-electron approximation (the Drude model) in describing the electronic properties of simple metals¹¹. When disorder is introduced, with the magnitude of disorder potential comparable to the bandwidth, multiple scattering causes the electronic states near the Fermi energy (E_F) to become localized and thereby causes a transition from metal to insulator¹². Strong disorder transforms the Fermi liquid into a ‘Fermi glass’¹³.

The disorder in metallic polymers arises from a combination of molecular-scale disorder^{2,3} and structural inhomogeneities at mesoscopic length scales^{4,6}. Even the most highly conducting polymers have been described as disordered metals near the metal–insulator transition^{2,3}. They have a finite density of states at E_F as verified by measurements of the Pauli spin susceptibility¹⁴, the linear term in the temperature dependence of the thermoelectric power¹⁵, and the linear term in the temperature dependent heat capacity¹⁶. Despite the finite density of states at E_F , the transport properties in the ‘metallic state’ as reported in the literature do not exhibit the traditional signatures of conventional metals. The d.c. conductivity (σ_{dc}) is thermally activated (phonon assisted), decreasing as the temperature is lowered^{4,17–19}, and $\sigma(\omega)$ is not consistent with Drude behaviour in the infrared^{20–24}. In the best cases, resistivity as a function of temperature, $\rho(T)$, is nearly constant, with a weak minimum below

room temperature^{2,3}. Thus, before this report, truly metallic transport has never been observed in conducting polymers.

Polyaniline (PANI) samples were prepared by self-stabilized dispersion polymerization (SSDP) as described in detail elsewhere⁸. In contrast to conventional homogeneous/dispersion polymerization using an aqueous medium containing aniline, acid and oxidant, this new polymerization was carried out in a heterogeneous biphasic (organic and aqueous) mixture without any stabilizers (Fig. 1). As the anilinium hydrochloride monomer is like a surfactant, with a polar hydrophilic part and an organic hydrophobic part, the monomers and growing polymer chains act as interfacial stabilizers, resulting in excellent dispersion of the organic phase inside the aqueous reaction medium. The organic phase tends to separate the aniline monomers and grown PANI chains from the reactive ends of the chain in the aqueous phase, thereby suppressing undesirable side reactions (such as *ortho*-coupling or Michael reductive additions). The reaction mixture was filtered, washed, deprotonated to the emeraldine base (EB) form, and then dried⁸. This synthesis produced high quality undoped PANI-EB samples with a low density of structural defects, as demonstrated in previous studies⁸.

Free-standing films of emeraldine salt were prepared by doping PANI-EB with camphor sulphonic acid (CSA), giving PANI-CSA, and cast from solution in *meta*-cresol. Typical sample thicknesses were about 30 μm . Figure 2 shows the temperature dependence of ρ for a series of free-standing films of PANI-CSA. For comparison, we

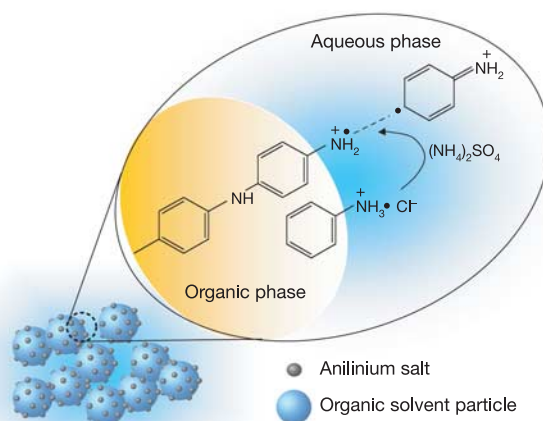


Figure 1 | Diagram of the self-stabilized dispersion polymerization of polyaniline. The polymerization is carried out in a heterogeneous biphasic (organic and aqueous) medium in order to prevent undesirable side reactions. The anilinium hydrochloride monomers locate on the surface of the organic solvent component and the radical is generated in the aqueous phase at the beginning of polymerization.

¹Department of Physics, Pusan National University, Busan 609-735, Korea. ²Center for Polymers and Organic Solids, University of California at Santa Barbara, Santa Barbara, California 93106-5090, USA. ³Department of Molecular Science and Technology, Ajou University, Suwon 442-749, Korea.

*These authors contributed equally to this work.

also plot $\rho(T)$ for the PANI-CSA that was prepared by the standard polymerization method ($\sigma_{dc}(300\text{ K}) \approx 270\text{ S cm}^{-1}$, and resistivity ratio $\rho_r = \rho(5\text{ K})/\rho(300\text{ K}) \approx 2.6$). Although the synthesis conditions are nominally identical for all samples, details of the film preparation and processing lead to differences in the sample quality, and thereby to changes in the electrical properties, as can be seen in Fig. 2. The SSDP samples exhibit significantly enhanced transport properties compared to those of conventional samples. As $\rho(300\text{ K})$ decreases, $\rho(T)$ remains progressively lower (σ_{dc} is higher) over the entire temperature range, the minimum in $\rho(T)$ shifts to lower temperatures, and the low temperature up-turn becomes less pronounced. For the highest conductivity samples, S1 and S2 (with respective $\sigma_{dc}(300\text{ K})$ values of $\sim 1,300\text{ S cm}^{-1}$ and $1,100\text{ S cm}^{-1}$), $\rho(T)$ exhibits a positive temperature dependence ($d\rho/dT > 0$) down to our low temperature measurement limit of 5 K with $\rho_r \approx 0.4$ for S1 (10 K for S2), as shown in Fig. 2 inset. Measurements of $\rho(T)$ below 5 K are under way for the highest conductivity samples.

The metallic nature of SSDP samples was confirmed by measurements of the reflectance, $R(\omega)$, at room temperature (Fig. 3a). Reflectance measurements were made of samples taken from all the synthesis batches. The data obtained from S1 and S2 (Fig. 2) were almost indistinguishable; the data shown in Fig. 3a were obtained from S2. In the far-infrared, $R(\omega) > 0.8$; $R(\omega)$ drops throughout the mid-infrared to a minimum at 1.4 eV, indicative of the free-carrier plasma resonance characteristic of a metal. Above the 1.4 eV minimum, $R(\omega)$ shows weak structure around 2.7 eV associated with the interband transition well known in doped PANI (and associated with the characteristic green colour). As shown in the inset of Fig. 3a, the Hagen-Rubens (H-R) approximation provides an excellent fit to $R(\omega)$ in the far-infrared.

The weak structure in the mid-infrared (1,000–1,500 cm^{-1}) arises from the infrared-active vibration (IRAV) modes characteristics of conducting polymers. Note, however, that the IRAV modes of these highest conductivity samples are particularly weak compared to

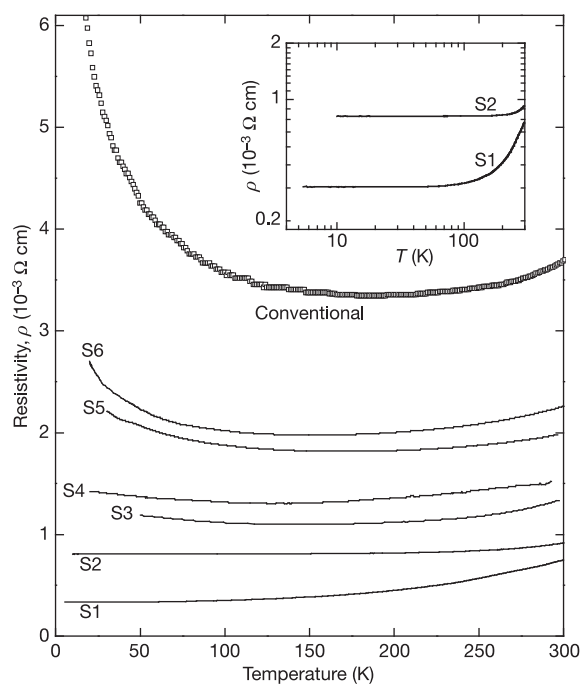


Figure 2 | Temperature dependence of resistivity, $\rho(T)$. Data from a conventional PANI-CSA film are shown for comparison. The SSDP PANI-CSA samples exhibit significantly enhanced conductivities compared with that obtained from the conventional sample. For more highly conducting samples (S5 $\rightarrow$ S1), the resistivity minimum weakens and shifts down to lower temperature and eventually disappears in the S1 and S2 samples (as shown more clearly in the inset).

those reported earlier for ‘metallic’ polyaniline²⁰. The IRAV modes in conducting polymers are characteristic of localized states, either self-localized states associated with solitons and polarons or localized states induced by disorder. When charges occupy such localized states, they move in response to the electric field of the infrared radiation and indirectly drive the amplitude modes, causing the Raman-active modes to become infrared-active. In the metallic regime, however, where the Fermi velocity is much greater than sound velocity, we do not expect to observe resonant IRAV modes. Thus, the reduced oscillator strength of the IRAV modes is consistent with the $\rho(T)$ data and implies that the metallic PANI-CSA characterized in Figs 2 and 3 is well beyond the insulator-to-metal transition.

Although the spectral features in $R(\omega)$ are similar to those of conventional PANI-CSA, the higher quality of the present samples is

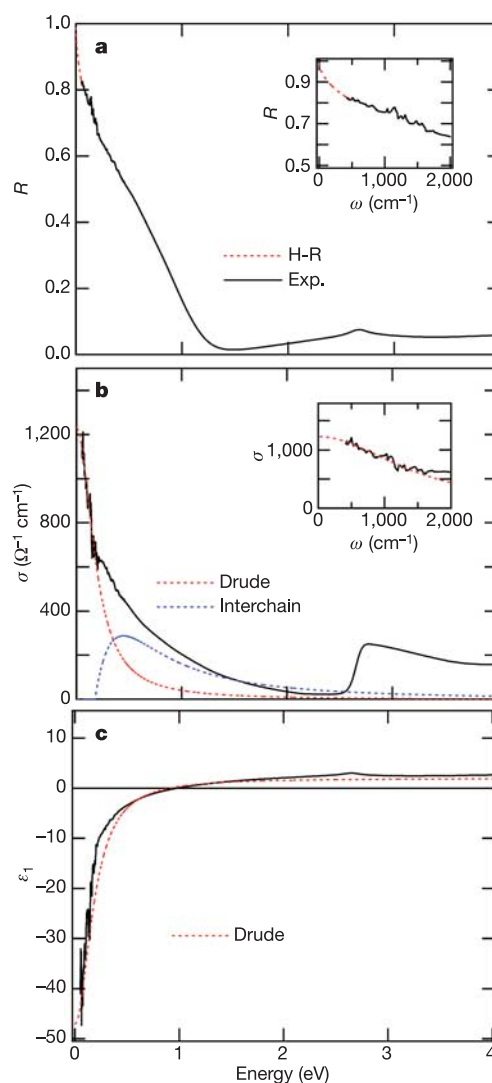


Figure 3 | Optical spectra of high-conductivity PANI-CSA. **a**, Reflectance spectra, $R(\omega)$ (solid line shown as ‘Exp.’). The Hagen-Rubens approximation (H-R, red dotted line) provides an excellent fit to $R(\omega)$ in the far-infrared. The inset shows $R(\omega)$ below 2,000 cm^{-1} with an expanded scale. **b**, Frequency dependent conductivity, $\sigma(\omega)$ (solid line), as obtained from Kramers-Kronig analysis of $R(\omega)$. The red dotted line represents the theoretical fit by the Drude model, $\sigma_D(\omega) = (\omega_p^2\tau/4\pi)(1 + \omega^2\tau^2)^{-1}$, with $\omega_p = 10,500\text{ cm}^{-1}$ and $\tau = 2.2 \times 10^{-14}\text{ s}$, whereas the blue dotted line corresponds to the interchain contribution to $\sigma(\omega)$. **c**, The real part of the dielectric function, $\epsilon_1(\omega)$ (solid curve). The red dotted line represents the theoretical fit by the Drude model, $\epsilon_{1D}(\omega) = \epsilon_\infty - \omega_p^2\tau^2(1 + \omega^2\tau^2)^{-1}$, using the same parameters as in Fig. 3b.

evident from the optical conductivity, $\sigma(\omega)$, and the real part of the dielectric function, $\varepsilon_1(\omega)$, as obtained through the Kramers-Kronig (K-K) analysis. Figure 3b shows $\sigma(\omega)$ for sample S2. In contrast with the results obtained from conventional conducting polymers^{20–24}, in which $\sigma(\omega)$ decreases below $\omega = 2,500 \text{ cm}^{-1}$, the PANI-CSA data in Fig. 3b follow the typical Drude frequency dependence at low frequencies (below $2,000 \text{ cm}^{-1}$) and approaches the measured $\sigma_{\text{dc}}(300 \text{ K})$. The excellent fit to $\sigma(\omega)$ of the classical Drude model, $\sigma_{\text{D}}(\omega) = (\omega_p^2 \tau / 4\pi)(1 + \omega^2 \tau^2)^{-1}$, with the plasma frequency $\omega_p = 10,500 \text{ cm}^{-1}$ ($\sim 1.3 \text{ eV}$) and relaxation time $\tau = 2.2 \times 10^{-14} \text{ s}$ ($1/\tau = 1,500 \text{ cm}^{-1} = 0.19 \text{ eV}$), is shown in Fig. 3b inset. For $\omega > 2,000 \text{ cm}^{-1}$, the interchain contribution starts to dominate $\sigma(\omega)$, consistent with previous studies²⁵.

The $\varepsilon_1(\omega)$ spectrum is also consistent with the Drude model and true metallic behaviour. The fit to $\varepsilon_1(\omega)$ of the Drude model, $\varepsilon_{1\text{D}}(\omega) = \varepsilon_\infty - [\omega_p^2 \tau^2 / (1 + \omega^2 \tau^2)]$, with the same values of ω_p and τ as in $\sigma_{\text{D}}(\omega)$ and with the high frequency dielectric constant $\varepsilon_\infty \approx 2$, shows again good agreement with the data (Fig. 3c). In the data obtained from earlier reflectance studies of conducting polymers²⁰, ε_1 crosses zero and becomes negative at the screened plasma frequency given by $\Omega_p = \omega_p / (\varepsilon_\infty)^{1/2}$, as expected for a typical metal. However, in contrast to simple Drude behaviour, $\varepsilon_1(\omega)$ crosses zero again to positive values around $\omega \approx 2,500 \text{ cm}^{-1}$ and remains positive down to $\sim 20 \text{ cm}^{-1}$. The $\varepsilon_1(\omega)$ data for PANI-CSA reported here, however, cross zero at $\omega = 7,800 \text{ cm}^{-1}$ ($= 0.97 \text{ eV}$) and remain negative at all frequencies below $7,800 \text{ cm}^{-1}$, again consistent with the classic Drude model.

The parameters obtained from the Drude fit are reasonable and consistent with other experimental results. The plasma frequency $\omega_p \approx 1.3 \text{ eV}$ is in good agreement with the $R(\omega)$ minimum around 1.4 eV , and yields $\Omega_p \approx 1 \text{ eV}$ with $\varepsilon_\infty \approx 2$, which is exactly consistent with the zero-crossing frequency in $\varepsilon_1(\omega)$. Using these parameters,

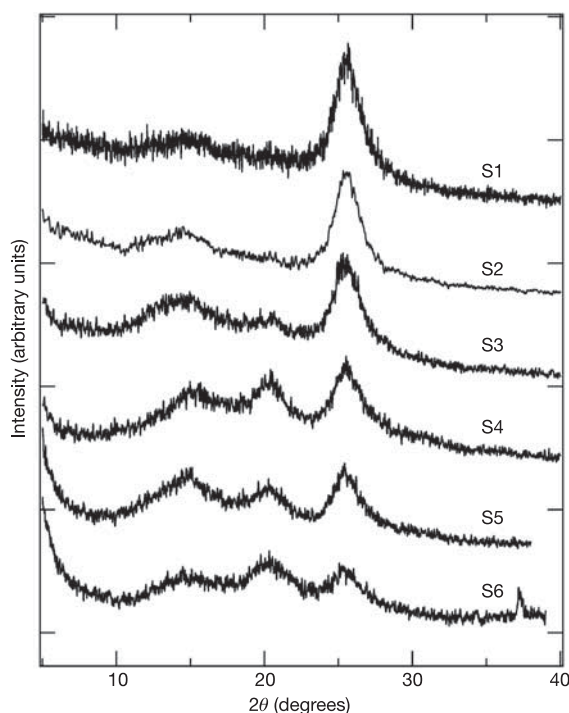


Figure 4 | X-ray diffraction patterns of the SSDP PANI-CSA samples. The low conductivity sample (S6) is almost identical with that of the conventional PANI. For the more metallic samples (S5 → S1), the broad background from the amorphous phase disappears and the crystalline peaks (at 15° , 20° and 25°) change gradually. For the best metallic samples (S1 and S2), the peak at 25° dominates (the background and the peaks at 15° and 20° have nearly disappeared).

we estimate the d.c. conductivity to be $\sigma_{\text{D}}(\omega = 0) = \omega_p^2 \tau / 4\pi \approx 1,200 \text{ S cm}^{-1}$, in good agreement with the measured $\sigma_{\text{dc}}(300 \text{ K}) \approx 1,100 \text{ S cm}^{-1}$. Moreover, from the relation $\omega_p = (4\pi n e^2 / m^*)^{1/2}$ where n is the charge carrier density and m^* is the effective mass, we obtain $n = 2.5 \times 10^{21} \text{ cm}^{-3}$ with the reasonable assumption of $m^* \approx 2m_e$ (m_e , free electron mass) for this class of materials^{20,21}. This result is consistent with that obtained from Hall measurements on this sample; the Hall effect data give $n_{\text{Hall}} = 2.1 \times 10^{21} \text{ cm}^{-3}$. In the previous studies on conventional conducting polymers^{20,21}, the proper description of the charge dynamics was achieved only after ‘modifying’ the Drude model in the context of weak localization. However, the excellent agreement established here demonstrates that high conductivity PANI-CSA can be successfully described by the simple Drude model without incorporating any contributions from disorder-induced localization theory.

We attribute the present results to the improved molecular structure of PANI-CSA prepared by SSDP. X-ray diffraction (XRD) patterns of films (taken in the reflection geometry) are shown in Fig. 4. For the lowest conductivity sample (S6), the diffraction pattern is almost identical with that reported earlier for PANI-CSA, consisting of three sharp peaks superimposed on a broad scattering background indicative of crystalline regions dispersed in an amorphous medium^{26,27}. However, for the more metallic samples (S5 → S1) the broad background decreases, and the relative intensity of the three peaks (at 15° , 20° and 25°) changes: the peaks at 15° and 20° decrease and become nearly indistinguishable from the background, whereas the peak at 25° increases and sharpens. As a consequence, for the highest conductivity samples with $d\rho(T)/dT > 0$, the diffraction pattern is dominated by the peak at 25° . The d -spacing ($\sim 3.5 \text{ \AA}$) associated with the diffraction peak at 25° corresponds to the face-to-face interchain stacking distance between phenyl rings. Thus, the increase in the intensity of the 25° peak along with the decrease in background intensity implies improved π – π interchain stacking. This suggests a more planar chain conformation with reduced torsion angles between the phenyl ring and the plane of the backbone, resulting in elongation of the effective conjugation length. These initial structural data are qualitatively consistent with improved carrier transport and a longer mean free path.

The achievement of metallic samples well on the metallic side of the insulator–metal transition for PANI-CSA implies that the promise of high performance semiconducting and metallic polymers could be realized by improving the material quality of known polymer systems. The utilization of such high performance semiconducting and metallic polymers in ‘plastic electronics’ would significantly extend the range of possible applications.

METHODS

Polyaniline samples were prepared at the Department of Molecular Science and Technology, Ajou University. All measurements ($\rho(T)$ versus T , optical reflectance, and XRD) were carried out in the Physics Department at Pusan National University. For the $\rho(T)$ versus T measurements, both the standard four-probe method and the Van der Pauw method were used in a closed cycle cryogenic system. A sapphire (Al_2O_3) plate was used to thermally anchor the samples to the Cu cold fingers; the samples were mounted on the sapphire plate using a GE-varnish adhesive. Gold wires in a four-probe configuration were connected parallel to the sample surface. Electrical contacts were made with conducting graphite adhesive. The current was applied using a Keithley 237 current source, while the voltage drop was measured using a Keithley 181 nanovoltmeter.

Each of the data sets in Fig. 2 was obtained from measurements of three independent samples from material from a single synthesis. Two samples were measured using the four-probe method, while a third sample was measured using the Van der Pauw method. The data from all three measurements were consistent and in quantitative agreement within experimental error. The highest conductivity data (S1) were obtained from three different samples taken from material from a single synthesis run. Materials with $\sigma_{\text{dc}}(300 \text{ K}) \approx 1,100 \text{ S cm}^{-1}$ (or less) were obtained from several independent synthesis runs.

Optical reflectance spectra were measured between 0.04 and 6.2 eV using two different spectrometers. Infrared reflectance in the range 0.04 – 0.5 eV was measured with a Mattson 5000 Fourier transform interferometer (FTIR),

while a Varian Cary 5E spectrophotometer covered the spectral range 0.4–6.2 eV. A gold mirror was used as the reference in the infrared range, and an aluminium mirror was used in the visible-ultraviolet range.

XRD experiments were performed using a conventional X-ray scanning diffractometer (Rigaku X-ray). The X-ray radiation source was Cu-K α radiation ($\lambda = 1.5418 \text{ \AA}$), and the scattered radiation diffractograms were collected over the range $2\theta \approx 5\text{--}40^\circ$ in a reflection geometry.

Received 2 May 2005; accepted 28 February 2006.

- Heeger, A. J. Nobel Lecture: Semiconducting and metallic polymers: The fourth generation of polymeric materials. *Rev. Mod. Phys.* **73**, 681–700 (2001).
- Menon, R., Yoon, C. O., Moses, D. & Heeger, A. J. in *Handbook of Conducting Polymers* 2nd edn (eds Skotheim, T. A., Elsenbaumer, R. L. & Reynolds, J. R.) 27–84 (Marcel Dekker, New York, 1998).
- Heeger, A. J. The critical regime of the metal-insulator transition in conducting polymers: experimental studies. *Phys. Scr.* **T102**, 30–35 (2002).
- Kaiser, A. B. Systematic conductivity behaviour in conducting polymers: effects of heterogeneous disorder. *Adv. Mater.* **13**, 927–941 (2001).
- Kiebooms, R., Menon, R. & Lee, K. in *Handbook of Advanced Electronic and Photonic Materials and Devices* Vol. 8 (ed. Nalwa, H. S.) 1–102 (Academic, San Diego, 2001).
- Kohlman, R. S., Joo, J. & Epstein, A. J. in *Physical Properties of Polymers Handbook* (ed. Mark, J. E.) 453–478 (AIP, New York, 1996).
- Lee, K. in *Encyclopedia of Nanoscience and Nanotechnology* Vol. 5 (ed. Nalwa, H. S.) 537–549 (American Scientific Publishers, San Diego, 2004).
- Lee, S.-H., Lee, D.-H., Lee, K. & Lee, C.-W. High-performance polyaniline prepared via polymerization in a self-stabilized dispersion. *Adv. Funct. Mater.* **15**, 1495–1500 (2005).
- Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Saunders College, New York, 1976).
- Landau, L. D. The theory of a Fermi liquid. *Sov. Phys. JETP* **3**, 920–925 (1957).
- Pines, I. D. & Nozieres, P. *The Theory of Quantum Liquids* (Benjamin, Menlo Park, 1966).
- Mott, N. F. *Metal-insulator Transitions* (Taylor & Francis, London, 1990).
- Anderson, P. W. The Fermi glass: theory and experiment. *Comments Solid State Phys.* **2**, 193–198 (1970).
- Fite, C., Cao, Y. & Heeger, A. J. Magnetic susceptibility of one-dimensional metallic chains in solution. *Solid State Commun.* **73**, 607–609 (1990).
- Park, Y. W., Choi, E. S. & Suh, D. S. Metallic temperature dependence of resistivity in perchlorate doped polyacetylene. *Synth. Met.* **96**, 81–86 (1998).
- Moses, D., Denenstien, A., Pron, A., Heeger, A. J. & MacDiarmid, A. G. Specific heats of pure and doped polyacetylene. *Solid State Commun.* **36**, 219–224 (1980).
- Reghu, M., Cao, Y., Moses, D. & Heeger, A. J. Counterion-induced processibility of polyaniline: transport at the metal-insulator boundary. *Phys. Rev. B* **47**, 1758–1764 (1993).
- Yoon, C. O., Reghu, M., Moses, D. & Heeger, A. J. Transport near the metal-insulator transition: polypyrrole doped with PF₆. *Phys. Rev. B* **49**, 10851–10863 (1994).
- Wang, Z. H., Scherr, E. M., MacDiarmid, A. G. & Epstein, A. J. Transport and EPR studies of polyaniline: a quasi one-dimensional conductor with three dimensional “metallic” states. *Phys. Rev. B* **45**, 4190–4202 (1992).
- Lee, K., Heeger, A. J. & Cao, Y. Reflectance of polyaniline protonated with camphor sulfonic acid: disordered metal on the metal-insulator boundary. *Phys. Rev. B* **48**, 14884–14891 (1993).
- Lee, K. *et al.* Nature of the metallic state in conducting polypyrrole. *Adv. Mater.* **10**, 456–459 (1998).
- Kohlman, R. S. *et al.* Limit for metallic conductivity in conducting polymers. *Phys. Rev. Lett.* **78**, 3915–3918 (1995).
- Chapman, B. *et al.* Low-energy conductivity of PF₆-doped polypyrrole. *Phys. Rev. B* **60**, 13479–13483 (1999).
- Tzamalīs, G., Zaidi, N. A., Homes, C. C. & Monkman, A. P. Doping-dependent studies of the Anderson-Mott localization in polyaniline at the metal-insulator boundary. *Phys. Rev. B* **66**, 085202 (2002).
- Lee, K. & Heeger, A. J. Optical investigation of intra and interchain charge dynamics in conducting polymers. *Synth. Met.* **128**, 279–282 (2002).
- MacDiarmid, A. G. & Epstein, A. J. Secondary doping in polyaniline. *Synth. Met.* **69**, 85–92 (1995).
- Yan, H., Ohta, T. & Toshima, N. Stretched polyaniline films doped by ($\pm$)-10-camphorsulfonic acid: anisotropy and improvement of thermoelectric properties. *Macromol. Mater. Eng.* **286**, 139–142 (2001).

Acknowledgements We thank K. Bechgaard for discussions. K.L. was supported at PNU by the National Program for Nanoscience and Technology of the Korea Science and Engineering Foundation. S.-H.L. acknowledges financial support at AU from the Korea Science and Engineering Foundation through the Hyper Structured Organic Materials Research Center in Seoul National University. Work at UCSB was supported by the National Science Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.L. (kwhlee@pusan.ac.kr) and S.-H.L. (hyja@ajou.ac.kr).

Towards molecular electronics with large-area molecular junctions

Hylke B. Akkerman¹, Paul W. M. Blom¹, Dago M. de Leeuw² & Bert de Boer¹

Electronic transport through single molecules has been studied extensively by academic^{1–8} and industrial^{9,10} research groups. Discrete tunnel junctions, or molecular diodes, have been reported using scanning probes^{11,12}, break junctions^{13,14}, metallic crossbars⁶ and nanopores^{8,15}. For technological applications, molecular tunnel junctions must be reliable, stable and reproducible. The conductance per molecule, however, typically varies by many orders of magnitude⁵. Self-assembled monolayers (SAMs) may offer a promising route to the fabrication of reliable devices, and charge transport through SAMs of alkanethiols within nanopores is well understood, with non-resonant tunnelling dominating the transport mechanism⁸. Unfortunately, electrical shorts in SAMs are often formed upon vapour deposition of the top electrode^{16–18}, which limits the diameter of the nanopore diodes to about 45 nm. Here we demonstrate a method to manufacture molecular junctions with diameters up to 100 μm with high yields (>95 per cent). The junctions show excellent stability and reproducibility, and the conductance per unit area is similar to that obtained for benchmark nanopore diodes. Our technique involves processing the molecular junctions in the holes of a lithographically patterned photoresist, and then inserting a conducting polymer interlayer between the SAM and the metal top electrode. This simple approach is potentially low-cost and could pave the way for practical molecular electronics.

The theoretical prediction of unipolar molecular diodes⁴ has led to a global effort to experimentally verify such devices. The envisaged application is molecular electronics, which could potentially solve fundamental scaling limits in complementary metal oxide semiconductor (CMOS) integrated circuits. Electronic transport in single molecules has been investigated with break junctions and a variety of scanning probe techniques. The conductance per molecule, however, differs by orders of magnitude. This discrepancy is probably due to the presence of an additional tunnelling gap in scanning probe techniques, to uncertainties in the number of molecules measured, or to poorly defined device surface areas (ref. 5 and references therein). Well-defined junctions can only be realized when the molecules are sandwiched between two electrodes. To ensure that electrical transport is dominated by the molecules and will therefore scale with the device area, an ordered monolayer is required. Typically, SAMs of alkanethiols on noble metals such as gold are used (ref. 19, and ref. 20 and references therein). Densely packed mono-domains of up to several hundred square nanometres can easily be formed when the chain length exceeds ten units (ref. 20 and references therein, and ref. 21).

Metal/SAM/metal nanojunctions about 45 nm in diameter have been fabricated using electron beam lithography and suspended silicon nitride membranes⁸. The conductance through the SAM of alkane thiols in the nanopores is independent of temperature and exponentially dependent on the length of the molecule⁸, which

implies that the electrical transport mechanism is non-resonant through-bond tunnelling. Therefore, the conductance per unit area through a SAM of alkane(di)thiols can be considered a benchmark for any novel technology related to molecular electronics.

The critical step in the fabrication of molecular junctions is applying the metal top electrode. Vapour-deposited noble metals create shorts in the SAMs due to filamentary growth^{16–18}, while reactive metals such as Ti destroy the SAM^{16,18,22}. Here we report a technology that overcomes this difficulty and produces reliable, stable and reproducible molecular junctions. We prevent shorts by using a layer of highly conducting polymer between the SAM and the vapour-deposited metal top electrode. The molecular junctions are processed in vertical interconnects (or via holes) of photolithographically patterned photoresist, which eliminates parasitic currents and protects the junction from the environment.

The processing of molecular junctions of large area with diameters of 10–100 μm is schematically depicted in Fig. 1 (see Supplementary Information for details). On a 4-inch silicon wafer with a thermally grown oxide, a 1 nm chromium adhesion layer and a 40 nm gold bottom contact are vapour-deposited (Fig. 1a) with a typical root-mean-square roughness of about 0.5 nm for a 1 μm^2 area (see Supplementary Information). Photoresist is spin-coated and holes

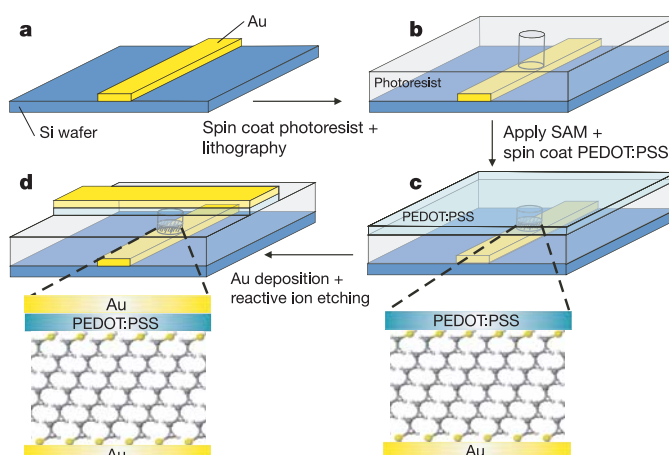


Figure 1 | Processing steps of a large-area molecular junction. **a**, Gold electrodes are vapour-deposited on a silicon wafer and a photoresist is spin-coated. **b**, Holes are photolithographically defined in the photoresist. **c**, An alkane dithiol SAM is sandwiched between a gold bottom electrode and the highly conductive polymer PEDOT:PSS as a top electrode. **d**, The junction is completed by vapour-deposition of gold through a shadow mask, which acts as a self-aligned etching mask during reactive ion etching of the PEDOT:PSS. The dimensions for these large-area molecular diodes range from 10 to 100 μm in diameter.

¹Materials Science Centre^{Plus}, University of Groningen, Nijenborgh 4, NL-9747 AG, Groningen, The Netherlands. ²Philips Research Laboratories, High Tech Campus 4, NL-5656 AA, Eindhoven, The Netherlands.

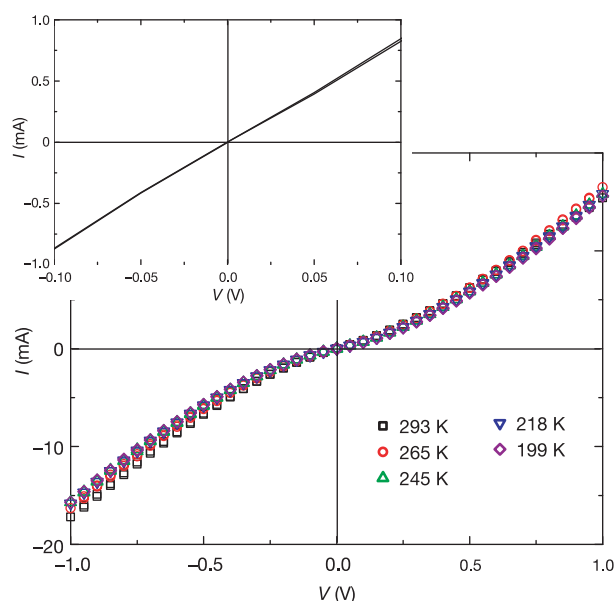


Figure 2 | *I*-*V* measurements at various temperatures of decanedithiol SAM on a device of diameter 100 μm . The *I*-*V* characteristics are temperature independent (199–293 K), showing tunnelling as the dominant transport mechanism. Up and down scans are plotted and no hysteresis is observed. Inset, at low bias, a linear *I*-*V* relationship is obtained, as expected for tunnel junctions.

of diameter 10–100 μm are produced by standard photolithography (Fig. 1b). The substrate is submerged in a freshly prepared 3×10^{-3} M solution of the alkane dithiol in ethanol for a minimum of 36 h. After the self-assembly of the alkane dithiolate on the gold bottom electrode, a water-based suspension of poly(3,4-ethylenedioxythiophene) stabilized with poly(4-styrenesulphonic acid) (PEDOT:PSS) is spin-coated on top of the SAM, resulting in a layer thickness of about 90 nm (Fig. 1c). PEDOT:PSS is a commercially available, highly doped, conductive polymer with a conductivity of about 30 S cm^{-1} . Next, the top gold electrode is vapour-deposited through a shadow mask. The top electrode acts as a self-aligned etching mask during the removal of the exposed PEDOT:PSS using reactive ion etching (Fig. 1d and Supplementary Information). The yield of functional molecular junctions is over 95%. Here we focus on SAMs of alkane dithiols. SAMs of alkane monothiol are more difficult to use in the processing of hydrophilic PEDOT:PSS owing to their hydrophobic CH_3 end groups. The monothiol-based molecular junctions, however, exhibit similar current versus applied voltage (*I*-*V*) characteristics, although with a larger standard deviation.

All *I*-*V* measurements were performed in vacuum to avoid the influence of water in PEDOT:PSS (see Supplementary Information). The *I*-*V* characteristic of a typical metal–insulator–metal junction 100 μm in diameter and based on decanedithiol is presented in Fig. 2. The current is dominated by the alkane dithiol monolayer. The current in junctions without the SAM is orders of magnitude larger. Figure 2 shows that at low bias the current increases linearly with applied bias (the low bias range is magnified in the inset of Fig. 2). This is expected for a metal–insulator–metal junction when the mean tunnel barrier height is larger than the applied voltage²³. Similarly, Fig. 2 shows the expected behaviour for a metal–insulator–metal junction at high bias, namely a current that increases exponentially with applied voltage²³. Furthermore, tunnelling is a temperature-independent process. The absence of any temperature dependence over the range from 199 to 293 K demonstrates that non-resonant tunnelling is the dominant transport mechanism⁸ (Fig. 2).

This conclusion is further substantiated by the dependence of the

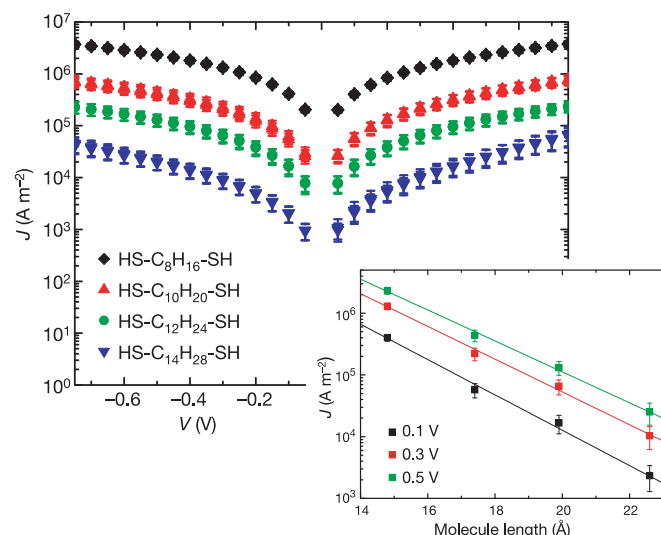


Figure 3 | The current density versus applied voltage for 1,8-octanedithiol, 1,10-decanedithiol, 1,12-dodecanedithiol and 1,14-tetradecanedithiol. The error bars give the standard deviation upon averaging over at least 17 devices. Both up and down scans are plotted. Inset, *J* (on a log scale) at 0.1, 0.3 and 0.5 V bias as a function of the molecular length. This indicates that the transport is by through-bond tunnelling.

current density on the barrier width. For molecular tunnel junctions, through-bond tunnelling is assumed (ref. 5 and references therein), which implies that the electrons do not tunnel through vacuum but through the lowest unoccupied molecular orbital (LUMO). This is because the mean tunnel barrier height from the Fermi level of the metal to the LUMO is smaller than the work function of the metal. For through-bond tunnelling the width of the tunnel barrier is then equal to the length of the molecule. In Fig. 3 the current density is plotted on a logarithmic scale as a function of the applied voltage for various alkane dithiols: octanedithiol ($\text{HS-C}_8\text{H}_{16}\text{-SH}$), decanedithiol ($\text{HS-C}_{10}\text{H}_{20}\text{-SH}$), dodecanedithiol ($\text{HS-C}_{12}\text{H}_{24}\text{-SH}$) and tetradecanedithiol ($\text{HS-C}_{14}\text{H}_{28}\text{-SH}$). The synthesis of these alkane-dithiols is described in the Supplementary Information. No hysteresis is observed in the *J*-*V* characteristics during all the voltage sweeps. The error bars represent the standard deviation upon averaging over at least 17 devices. Figure 3 demonstrates that the current density decreases with increased length of the alkane dithiol molecule.

The current density at a bias of 0.1, 0.3 and 0.5 V is presented in the inset of Fig. 3 on a logarithmic scale as a function of the length of the molecule. Molecular lengths were calculated using ACD Labs software (see the Supplementary Information), and the linear fit through the data shows that *J* depends exponentially on the length of the alkane dithiols. This confirms that the currents are indeed specific for the molecules in the junctions. Furthermore, it demonstrates that non-resonant, through-bond tunnelling is the transport mechanism in these metal–insulator–metal junctions. Because current density $J \propto \exp(-\beta d)$ (where *d* is the barrier width; ref. 23), the tunnelling decay coefficient β can be deduced from the slope of this linear fit. We find β values of 0.66 ± 0.06 , 0.61 ± 0.05 , and $0.57 \pm 0.05 \text{ \AA}^{-1}$ at a bias of 0.1, 0.3 and 0.5 V, respectively. These values are in good agreement with those derived from analysis of the nanopore diodes by Wang *et al.*, who report values of β ranging from 0.83 to $0.72 \text{ \AA}^{-1}$ in the bias range from 0.1 to 1.0 V (ref. 8). The decrease of β with increasing bias is also as expected (ref. 5 and references therein). That the current density of our large-area molecular junctions exhibits a dependence on voltage and temperature identical to that of the nanopore devices, combined with the absence of any hysteresis, clearly demonstrates that PEDOT:PSS can be regarded as a non-interacting conductive electrode.

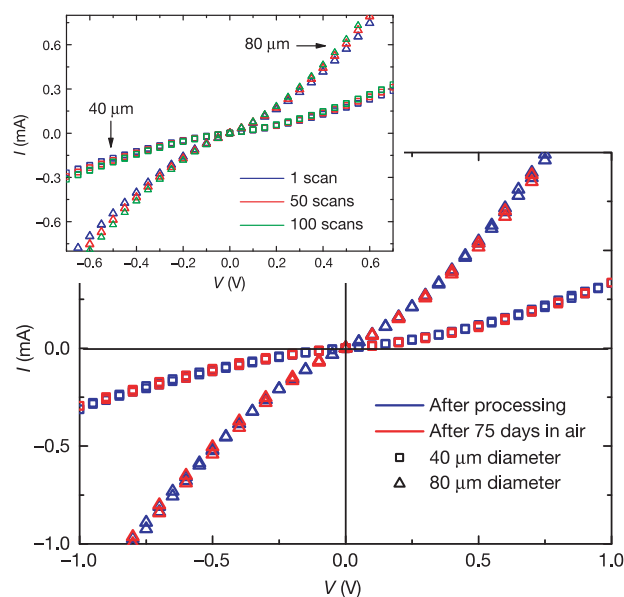


Figure 4 | Stability and operational lifetime of molecular junctions.

I-*V* characteristics of molecular junctions based on dodecane dithiol with diameters of 40 μm (squares) and 80 μm (triangles) measured both directly after processing (blue) and after storing under ambient conditions for 75 days (red). Identical *I*-*V* responses are obtained, which demonstrates the excellent stability of these molecular junctions. The inset shows three *I*-*V* measurements out of 100 consecutive sweeps of a molecular junction stored under ambient conditions for 75 days with diameters of 40 μm (squares) and 80 μm (triangles). No sign of hysteresis is observed.

For practical applications of molecular tunnel junctions, yield, reliability, operational stability and shelf-life are crucial. The yield of functional devices is larger than 95% even for molecular junctions with a diameter of 100 μm . The yield is presumably due to the large differences in surface tension between the hydrophilic PEDOT:PSS and the hydrophobic backbone of the alkane thiolates. This difference prevents intermixing and hence the formation of shorts. The reproducibility is remarkable and investigations into the junction stability are also promising. Figure 4 displays the current-voltage characteristics of molecular junctions based on dodecane dithiol with diameters of 40 μm and 80 μm both directly after fabrication and after storage under ambient conditions for 75 days. The data points fall on top of each other, showing that there is no sign of deterioration upon storage. The inset of Fig. 4 shows three *I*-*V* measurements taken from 100 consecutive forward and backward sweeps. No hysteresis is observed and there are no differences between consecutive sweeps. The excellent stability is presumably due to encapsulation of the diodes by photoresist, which prevents any interaction of the environment with the SAM.

To compare our molecular junctions to other data appearing in the literature^{5,8,12}, we normalized the current to the current per single molecule. A grafting density for alkanethiols^{5,20} of $4.6 \times 10^{18} \text{ m}^{-2}$ was assumed. As a typical example we use our data for dodecane-dithiol. The current per molecule at 0.2 V bias is plotted in Fig. 5 as a function of device area. Literature data are included for measurements on decanethiol in nanopore junctions and by scanning probe experiments, which covers a device area spanning ten orders of magnitude. Figure 5 shows good agreement between our data and the nanopore data⁸, with a difference in current per molecule of less than a factor of two. This difference is presumably due to the use of a monothiol in the nanopore junctions and a dithiol in our large-area molecular junctions. Hence, the tunnel barrier width in the nanopore experiments is slightly smaller, resulting in a higher current per molecule. The large scatter in the conducting probe and tunnelling microscope data is intrinsic and due to the presence of an additional

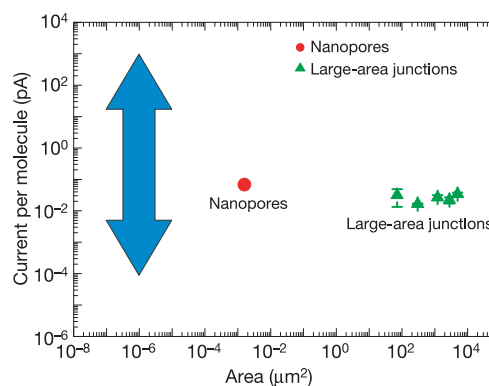


Figure 5 | Current normalized per single molecule for decane(di)thiols at 0.2 V bias versus the device area. Our large-area molecular junctions are presented by the green triangles. Error bars present the standard deviation. Values of current per molecule from the benchmark nanopore junctions are presented by the red circle. Perfect agreement is obtained despite a difference in device area of seven orders of magnitude. The blue arrow presents the intrinsically large scatter in current per molecule obtained from conducting probe atomic force microscopy and scanning tunnelling microscopy measurements.

tunnelling gap and uncertainties in the number of molecules measured. The good agreement between the current per molecule as measured in our large-area diodes and the nanopore junctions is remarkable considering the orders-of-magnitude difference in device area. In the nanopore junctions the current is measured through molecules that are perfectly ordered in a mono-domain, while in the large-area diodes the molecules are ordered in multi-domains. Grain boundaries in SAMs do not seem to impede charge transport.

In summary, we have demonstrated a technique for fabricating reliable molecular metal-insulator-metal junctions with unprecedented device diameters of up to 100 μm . The yield of these molecular junctions is close to 100%. Preliminary stability investigations reveal a shelf life of more than several months and no deterioration upon cycling. The current per molecule is similar to that obtained for benchmark nanopore diodes. Key ingredients are the use of a conducting polymer layer sandwiched between the SAM and the top electrode to prevent electrical shorts, and processing within lithographically defined vertical interconnects (or via holes) to prevent both parasitic currents and interaction between the environment and the SAM. The technique is simple, compatible with standard integrated circuit fabrication processes, and can be scaled up and extended to any molecule and any metal bottom electrode on which an ordered SAM can be formed.

Received 6 September 2005; accepted 2 March 2006.

- Reed, M. A., Zhou, C., Muller, C. J., Burgin, T. P. & Tour, J. M. Conductance of a molecular junction. *Science* **278**, 252–254 (1997).
- Joachim, C., Gimzewski, J. K. & Aviram, A. Electronics using hybrid-molecular and mono-molecular devices. *Nature* **408**, 541–548 (2000).
- Nitzan, A. & Ratner, M. A. Electron transport in molecular wire junctions. *Science* **300**, 1384–1389 (2003).
- Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
- Salomon, A. *et al.* Comparison of electronic transport measurements on organic molecules. *Adv. Mater.* **15**, 1881–1890 (2003).
- Szuchmacher Blum, A. *et al.* Molecularly inherent voltage-controlled conductance switching. *Nature Mater.* **4**, 167–172 (2005).
- Collier, C. P. *et al.* Electronically configurable molecular-based logic gates. *Science* **285**, 391–394 (1999).
- Wang, W., Lee, T. & Reed, M. A. Mechanism of electron conduction in self-assembled alkanethiol monolayer devices. *Phys. Rev. B* **68**, 035416 (2003).
- Zhitenev, N. B., Meng, H. & Bao, Z. Conductance of small molecular junctions. *Phys. Rev. Lett.* **88**, 226801 (2002).
- Chen, Y. *et al.* Nanoscale molecular-switch crossbar circuits. *Nanotechnology* **14**, 462–468 (2003).
- Cui, X. D. *et al.* Reproducible measurements of single-molecule conductivity. *Science* **294**, 571–574 (2001).

12. Engelkes, V. B. *et al.* Length-dependent transport in molecular junctions based on SAMs of alkanethiols and alkanedithiols: Effect of metal work function and applied bias on tunneling efficiency and contact resistance. *J. Am. Chem. Soc.* **126**, 14287–14296 (2004).
13. Reichert, J. *et al.* Driving current through single organic molecules. *Phys. Rev. Lett.* **88**, 176804 (2002).
14. Smit, R. H. M. *et al.* Measurement of the conductance of a hydrogen molecule. *Nature* **419**, 906–909 (2002).
15. Zhou, C. *et al.* Nanoscale metal/self-assembled monolayer/metal heterostructures. *Appl. Phys. Lett.* **71**, 611–613 (1997).
16. de Boer, B. *et al.* Metallic contact formation for molecular electronics: Interactions between vapor-deposited metals and self-assembled monolayers of conjugated mono- and dithiols. *Langmuir* **20**, 1539–1542 (2004).
17. Haick, H., Ghabboun, J. & Cahen, D. Pd versus Au as evaporated metal contacts to molecules. *Appl. Phys. Lett.* **86**, 042113 (2005).
18. Haynie, B. C. *et al.* Adventures in molecular electronics: how to attach wires to molecules. *Appl. Surf. Sci.* **203/204**, 433–436 (2003).
19. de Boer, B. *et al.* Synthesis and characterization of conjugated mono- and dithiol oligomers and characterization of their self-assembled monolayers. *Langmuir* **19**, 4272–4284 (2003).
20. Love, J. C. *et al.* Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem. Rev.* **105**, 1103–1169 (2005).
21. Porter, M. D. *et al.* Spontaneously organized molecular assemblies. 4. Structural characterization of *n*-alkyl thiol monolayers on gold by optical ellipsometry, infrared spectroscopy, and electrochemistry. *J. Am. Chem. Soc.* **109**, 3559–3568 (1987).
22. Chang, S.-C. *et al.* Investigation of a model molecular-electronic rectifier with an evaporated Ti-metal top contact. *Appl. Phys. Lett.* **83**, 3198–3200 (2003).
23. Simmons, J. G. Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* **34**, 1793–1803 (1963).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Mulder, T. C. T. Geuns, E. A. Meulenlamp, E. Cantatore and S. Bakker for their assistance, and the Materials Science Centre^{Plus} for financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.d.B. (b.de.boer@rug.nl).

Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing

Gabriel A. Vecchi¹, Brian J. Soden², Andrew T. Wittenberg¹, Isaac M. Held¹, Ants Leetmaa¹ & Matthew J. Harrison¹

Since the mid-nineteenth century the Earth's surface has warmed^{1–3}, and models indicate that human activities have caused part of the warming by altering the radiative balance of the atmosphere^{1,3}. Simple theories suggest that global warming will reduce the strength of the mean tropical atmospheric circulation^{4,5}. An important aspect of this tropical circulation is a large-scale zonal (east–west) overturning of air across the equatorial Pacific Ocean—driven by convection to the west and subsidence to the east—known as the Walker circulation⁶. Here we explore changes in tropical Pacific circulation since the mid-nineteenth century using observations and a suite of global climate model experiments. Observed Indo-Pacific sea level pressure reveals a weakening of the Walker circulation. The size of this trend is consistent with theoretical predictions, is accurately reproduced by climate model simulations and, within the climate models, is largely due to anthropogenic forcing. The climate model indicates that the weakened surface winds have altered the thermal structure and circulation of the tropical Pacific Ocean. These results support model projections of further weakening of tropical atmospheric circulation during the twenty-first century^{4,5,7}.

The Walker circulation is fundamental to climate throughout the globe: its variations are closely linked to those of the El Niño/Southern Oscillation⁶ and monsoonal circulations over adjacent continents⁸, and variations in its intensity and structure affect climate across the planet^{8,9}. The strength of equatorial Pacific zonal wind-

stress, associated with the Walker circulation, is critical to equatorial Pacific Ocean circulation¹⁰ and to biogeochemical processes¹¹.

Observational and modelling evidence indicates that since the mid-nineteenth century tropical sea surface temperatures (SSTs) have warmed by 0.5–0.6 °C (refs 1–3). Climate models predict a weakening of the atmospheric convective overturning in response to surface warming driven by increases in greenhouse gases^{4,5,7}. One expects this weakening to be manifest, in part, as a reduction in the zonal overturning of tropical air⁴, a large component of which is the Pacific Walker circulation⁶.

The weakening of tropical atmospheric overturning circulations in response to warming can be understood in terms of the energy and mass balance of the ascending branch of these circulations. As the surface warms, the water vapour concentration in the lower troposphere increases by roughly 7% per °C of surface warming^{12,13}, consistent with the Clausius–Clapeyron equation and fixed relative humidity. However, the rate of precipitation (which on long time-scales is limited by the rate of change of the radiative cooling of the troposphere) increases more slowly—approximately 2% per °C warming^{1,14,15}. The global-mean rate of precipitation must be balanced by the moisture transport from the atmospheric boundary layer to the free troposphere, which is a product of the boundary layer water vapour concentration and the exchange of air between the boundary layer and free troposphere. Thus, the differential rate of response to surface warming of water vapour and precipitation

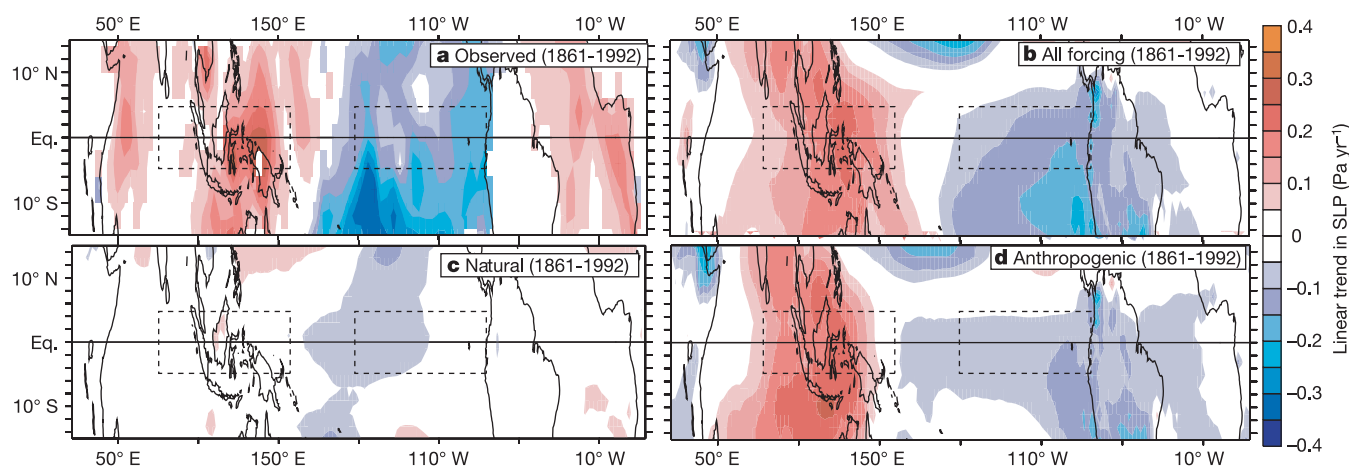


Figure 1 | Spatial pattern of observed and modelled sea level pressure linear trends. Linear trend of sea level pressure (SLP) from: **a**, Kaplan SLP reconstruction²⁹ (1861–1992), and ensemble-mean of GCM experiments (1861–1992) as follows; **b**, all-forcing (five-member mean), **c**, natural forcing

(three-member mean) and **d**, anthropogenic forcing (three-member mean). The trend averaged over the domain 15°S–15°N, 0°–360° is removed from each panel. Dashed rectangles indicate the regions used to define the large-scale Indo-Pacific SLP gradient index (Δ SLP).

¹NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey 08540-6649, USA. ²Rosenstiel School for Marine and Atmospheric Sciences, University of Miami, Miami, Florida 33149-1098, USA.

implies a weakening of the boundary layer/troposphere mass exchange of $\sim 5\%$ per $^{\circ}\text{C}$ warming⁴. Closely related arguments have been provided for a weakening of tropical overturning circulations based on the energy balance of the subsiding branch of these circulations⁵. On the basis of observed tropical warming since the mid-nineteenth century ($0.5\text{--}0.6^{\circ}\text{C}$), these theoretical arguments predict a 2.5–3% reduction in the strength of tropical atmospheric overturning circulation. Is there evidence of such a slowdown in the observational record?

We use historical observations of sea level pressure (SLP) to assess observed changes in the Walker circulation over the tropical Pacific; an index of the large-scale tropical Indo-Pacific SLP gradient (ΔSLP) serves as a proxy for the mean intensity of the Pacific Walker circulation (see Methods). Ensembles of global climate model (GCM) experiments—with different radiative forcings—serve to explore the origin of the observed circulation changes, and allow for the estimation of the statistical significance of the observed changes. The model used here is the US National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL) CM2.1 GCM^{16–18}, with three historical integration sets over the period 1861–2000: (1) a five-member ensemble including estimates of natural (solar variations, volcanoes) and anthropogenic (well-mixed greenhouse gases, ozone, direct aerosol forcing and land use) sources of climate change; (2) a three-member ensemble applying only natural forcing; and (3) a three-member ensemble applying only anthropogenic forcing (see Methods and Supplementary Information).

There have been spatially coherent patterns in observed trends of tropical SLP since the mid-nineteenth century (Fig. 1a), and similar patterns are evident in the five-member ensemble-mean of the historically forced GCM integrations (Fig. 1b). These trends indicate a reduced zonal SLP gradient, and thus a weakened zonal circulation, because the climatological pattern in the tropical Indo-Pacific has SLP larger in the east than in the west. The naturally forced GCM experiments are unable to recover these observed patterns in the SLP trends (Fig. 1c); however, the GCM recovers many of the principal

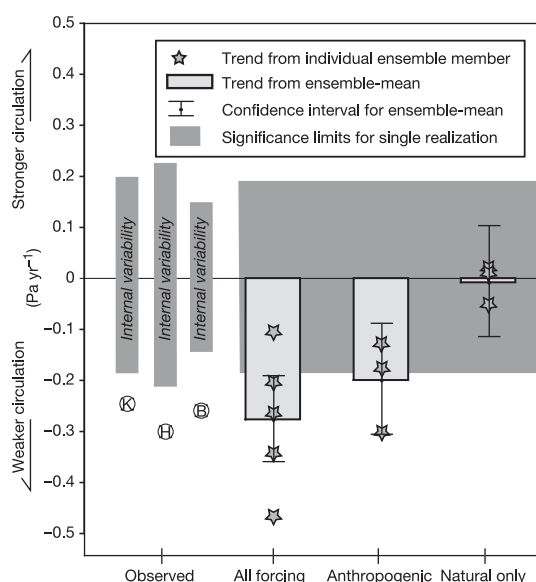


Figure 2 | Summary of the linear trends in SLP gradient across the Indo-Pacific (ΔSLP) from observations and the various GCM historical radiative forcing experiments. Circles indicate the trend value from each observational data set: K, Kaplan (1854–1992)²⁹; H, Hadley Centre (1871–1998)²⁸; and B, a blend of Hadley²⁸ and Kaplan²⁹, extended into 2005 using the NCEP gridded ship data (1854–2005)²⁷. Model trends are computed over the period 1861–2000. Confidence intervals are computed from a 2,000-year control experiment, at the two-sided $P = 0.05$ level.

observed SLP trend patterns using only anthropogenic forcing (Fig. 1d).

Trends computed from observed ΔSLP are inconsistent with those expected from the variability in the pre-industrial control simulation, and they are inconsistent with the trends from the ‘natural-forcing’ GCM ensemble experiment (Fig. 2). However, the trends in observed ΔSLP fall within the range of trends from the ‘all-forcing’ GCM ensemble, and within that of the ‘anthropogenic-forcing’ GCM ensemble (Fig. 2). Thus, within the framework of this GCM, a significant part of the observed reduction of ΔSLP since the mid-nineteenth century resulted from anthropogenic forcing. The trend computed from observed ΔSLP is also significantly distinct from that expected from the pre-industrial control experiments of all other models in the Intergovernmental Panel on Climate Change 4th Assessment Report (IPCC/AR4) archive (Supplementary Fig. 3). Most other models in the IPCC/AR4 archive also show a weakening of the equatorial Pacific pressure gradient, although CM2.1 shows the largest reduction (Supplementary Fig. 4).

Though we use linear trends to summarize the changes in tropical SLP since the mid-nineteenth century, these changes have not appeared as a smooth reduction (Fig. 3). Both the observational record and individual GCM ensemble members exhibit substantial decadal variability (Supplementary Fig. 1), arising from processes internal to the coupled system. The GCM decadal variability in ΔSLP is comparable to that in observations, even though the GCM interannual variability is overly energetic owing to a simulated El Niño variability that has too large an amplitude¹⁷. The strong decadal variability complicates the detection of a relatively small forced change even in multi-decadal records of ΔSLP ; on the basis of the 2,000-year control integration, a record shorter than 100–120 years is insufficient to detect the forced linear trend in ΔSLP from the GCM at $P = 0.05$ (Supplementary Fig. 2).

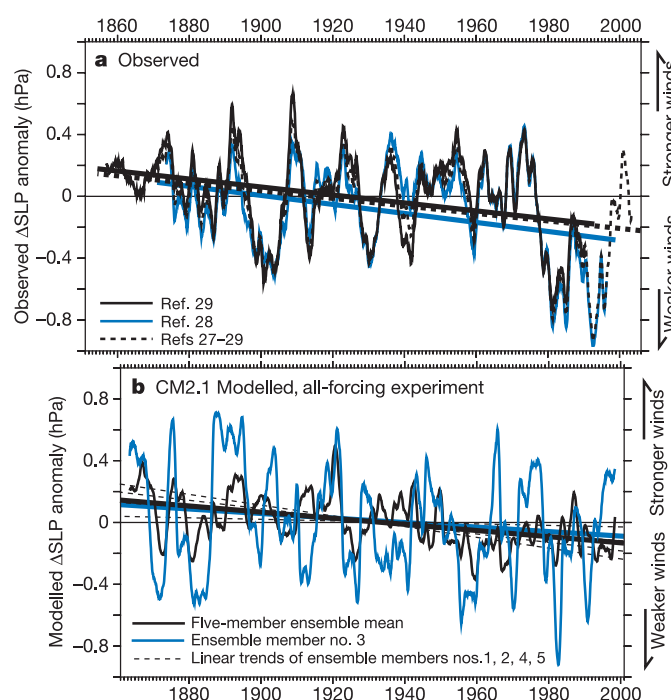


Figure 3 | Observed and modelled evolution of ΔSLP since the nineteenth century. Five-year running-mean ΔSLP from: **a**, observations (black from Kaplan²⁹, and blue from Hadley Centre²⁸; the record is extended through to 2005 with NCEP ship-based observations²⁷ (dashed line)); and **b**, GCM historical integrations (five-member ensemble mean in black, an illustrative ensemble member (number 3) in blue, dashed lines indicate linear trend of ensemble members 1, 2, 4 and 5). In both panels, the linear trends in ΔSLP are shown as thick lines, with shadings corresponding to each time-series.

In the 1970s there was a rapid reduction in observed Δ SLP; such rapid changes are also evident in other decades and in individual ensemble members. Fifty-year Δ SLP trends ending in the 1990s are significantly larger than those computed over the entire record, suggesting an amplifying trend in Δ SLP. However, in recent years the statistical significance of the amplification of the trend disappears; though nominally larger than the long-term trend, the 1956–2005 trend is not significantly different from the long-term trend at $P = 0.05$. Further work is required to determine if the larger recent trend since the 1950s is forced or a result of internal variability.

We estimate zonal wind stress across the equatorial Pacific using Δ SLP, following the method of ref. 19 (Fig. 4). Linear relationships with Δ SLP capture most of the interannual and longer timescale variability in equatorial Pacific zonal-mean easterlies ($\langle \tau^x \rangle$; zonal wind-stress averaged 120°E – 70°W , 5°S – 5°N ; easterlies are winds from the east). Δ SLP-reconstructed $\langle \tau^x \rangle$ recovers the principal extremes and transitions of the observed and modelled $\langle \tau^x \rangle$, including the long-term weakening of $\langle \tau^x \rangle$ in the model, and the various El Niño and La Niña events observed in the recent decades (Fig. 4, upper panels). The trend of the linear fit of $\langle \tau^x \rangle$ to observed Δ SLP represents a $\langle \tau^x \rangle$ reduction of $\sim 7\%$ since 1860; as surface stress is roughly proportional to the square of wind speed, this suggests a mean reduction in equatorial Pacific zonal wind of $\sim 3.5\%$, roughly consistent with the theoretical prediction.

Equatorial Pacific $\langle \tau^x \rangle$ is of critical importance to the large-scale oceanic circulation in the equatorial Pacific⁹. The GCM experiments indicate that since the mid-nineteenth century, the weakening of equatorial $\langle \tau^x \rangle$ has resulted in a weakening of surface equatorial currents, a vertical shift in sub-surface currents, and a reduction in the intensity and depth of equatorial upwelling (Supplementary Fig. 5). By bringing nutrient-rich waters close to the surface, equatorial upwelling exerts a strong control on biological activity

in the tropical Pacific¹⁰; its weakening and shoaling suggest a possible reduction of biological productivity under global warming.

The GCM experiments indicate a substantial shoaling of the western equatorial Pacific thermocline depth (Z_{tc}) since the mid-nineteenth century (Fig. 4, lower panels); the thermocline is the zone of rapid temperature change in a typical vertical oceanic temperature profile between the warm well-mixed surface layer and cold abyssal waters. Reduction of $\langle \tau^x \rangle$ affects both the east–west tilt of the equatorial Pacific thermocline and its mean depth²⁰ (Supplementary Fig. 6); these changes could affect the character of El Niño variability²¹. On seasonal to interannual timescales (timescales too short for the equatorial thermocline to come to equilibrium with the winds), a reduction in $\langle \tau^x \rangle$ has a strong impact on both western and eastern equatorial Pacific Z_{tc} . However, on timescales longer than that of equatorial adjustment, the impact of reductions in $\langle \tau^x \rangle$ is felt almost entirely by the western equatorial Pacific Z_{tc} (Fig. 4). Because Z_{tc} and SST are tightly coupled only in the eastern equatorial Pacific²², these long-term thermocline depth changes in the GCM are unlikely to affect SST directly. The observational record of equatorial Pacific thermocline slope and depth over the past 50 years is consistent with the recent reduction in the strength of the easterlies and the low-frequency relationship between Z_{tc} and $\langle \tau^x \rangle$ in the GCM.

Atmosphere–ocean conditions in the equatorial Pacific have changed since the mid-nineteenth century: there has been a significant slowdown of atmospheric circulation, which models indicate has driven a response in ocean circulation. The observed trend in the Pacific surface zonal SLP gradient is unlikely to be due to natural variability. However, much of the long-term trend is reproduced in model simulations that account for human impacts on the radiative budget of the planet, and is consistent with the changes expected from simple thermodynamic arguments⁴. The agreement between

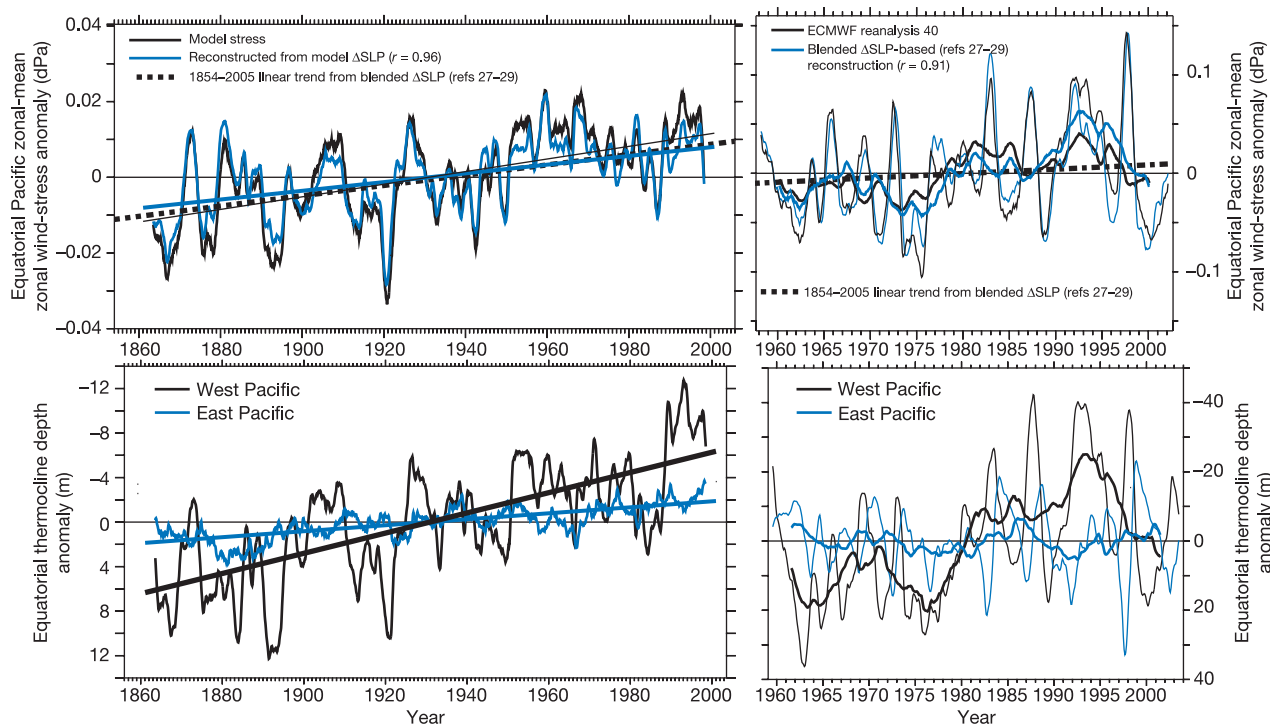


Figure 4 | Observed and modelled equatorial Pacific zonal-mean zonal wind-stress anomaly, $\langle \tau^x \rangle$, and equatorial thermocline depth anomaly, Z_{tc} . Upper panels: model/observed $\langle \tau^x \rangle$ and reconstruction using linear relation to Δ SLP; dashed line shows (1854–2005) trend in $\langle \tau^x \rangle$ reconstructed using blended Kaplan²⁹/Hadley²⁸/NCEP²⁷ Δ SLP. Lower panels: Z_{tc} in the western (black line, 2°S – 2°N , 140°E – 180°E) and eastern (blue line, 2°S – 2°N , 130°W – 90°W) equatorial Pacific. Left panels:

ensemble-mean all-forcing CM2.1 GCM experiment, showing five-year running mean. Right panels: five-year running mean (thick lines) and annual-mean (thin lines) observational estimates. Observed stress is from European Centre for Medium Range Weather Forecasting Reanalysis 40, observed Z_{tc} is from GFDL ocean data assimilation³⁰. Z_{tc} is the location of the maximum vertical temperature gradient. Note different scales in each panel.

the theoretical, observed and model-produced changes in strength of atmospheric circulation suggests increased confidence in the model-projected reduction in the strength of tropical circulation during the twenty-first century^{4,5,7}; on the basis of climate model simulations, this weakening may be of the order of 10% by the end of the twenty-first century^{4,7}.

METHODS

SLP data sets and calculations. Direct assessment of long-term changes in the strength of tropical circulation is problematic, because there have been changes in the methods used to make wind measurements at sea^{23,24}. Trends in tropical winds over recent decades are ambiguous, with some studies showing a strengthening²⁵ and others a weakening²⁶. SLP offers a proxy to recover changes in wind velocity, because of the dynamical connection between large-scale zonal gradients of SLP and zonal-mean zonal wind-stress¹⁹. Further, SLP measurements at sea have always been made using instruments, in contrast with measurements of surface wind, which have been instrument-based only in recent decades^{23,24}. Equatorial Pacific wind-stress estimated from the zonal gradients of SLP has shown a reduction over the period 1920–90¹⁹. Ship-based SLP measurements²⁷ have been used to build global gridded data sets of monthly SLP^{28,29} using knowledge of its global co-variability, and allow the exploration of changes from the mid-nineteenth century.

In this study two global reconstruction data sets of monthly SLP are used: (1) the Kaplan SLP data set²⁹ version 1 spanning the period 1854–1992, with data only over the ocean, and (2) the Hadley Centre SLP reconstruction²⁸ version 1 spanning the period 1871–1998, with data over both land and ocean. These data sets are extended into recent years using the gridded monthly ship-based SLP observations available from NOAA's National Center for Environmental Prediction (NCEP)²⁷. For the analyses presented here, the monthly long-term climatology of each data set is removed to compute SLP anomalies.

Using the GCM and observed SLP data, a large-scale tropical Indo-Pacific SLP gradient (Δ SLP) is computed from the difference in SLP averaged over the central/east Pacific (160°W–80°W, 5°S–5°N) and over the Indian Ocean/west Pacific (80°E–160°E, 5°S–5°N). The index is computed with SLP anomalies from monthly climatology; positive values indicate a strengthened Indo-Pacific SLP gradient. A least-squares linear trend in Δ SLP is used as a concise metric for the long-term changes in the strength of zonal circulation. As shown in ref. 19 a large-scale SLP gradient, like Δ SLP, provides a useful proxy for the mean intensity of Pacific zonal surface winds. For the observational record of surface winds available to us (1957–2003) and for the entire climate model record (1861–2000), Δ SLP provides a useful proxy for the strength of the mean zonal circulation over the equatorial Pacific Ocean (Fig. 4).

Climate models. Global climate models (GCMs) simulate the variations of, and interactions between, various elements of the climate system (ocean, atmosphere, cryosphere and land) forced by radiatively active naturally occurring and anthropogenic gases and aerosols. Internal climate variability can be isolated from that forced by changes to atmospheric composition through ensemble experiments, in which the same model physics and forcing fields are applied to different initial conditions. The model used here is the NOAA GFDL CM2.1 GCM^{16–18}, which uses estimated radiative forcings over the period 1861–2000. The principal historical integration set is a five-member ensemble including estimates of natural (solar variations, volcanoes), and anthropogenic (well-mixed greenhouse gases, ozone, direct aerosol forcing and land use) sources of climate change. Two additional sets of experiments isolate the effects of each set of forcing elements, by applying only natural or anthropogenic forcing; each consists of a three-member ensemble. Statistical significance estimates are computed from a 2,000-year control integration with invariant radiative conditions from the 1860s (see Supplementary Information).

Received 27 October 2005; accepted 22 March 2006.

1. Houghton, J., et al. *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, UK, 2001).
2. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**(D14), 4407, doi:10.1029/2002JD002670 (2003).
3. Knutson, T. R. et al. Assessment of twentieth-century regional surface

- temperature trends using the GFDL CM2 coupled models. *J. Clim.* (in the press).
4. Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* (in the press).
5. Knutson, T. R. & Manabe, S. Time-mean response over the tropical Pacific to increased CO₂ in a coupled ocean-atmosphere model. *J. Clim.* **8**, 2181–2199 (1995).
6. Julian, P. R. & Chervin, R. M. A study of the Southern Oscillation and the Walker Circulation. *Mon. Weath. Rev.* **106**, 1433–1451 (1978).
7. Tanaka, H. L., Ishizki, N. & Kitoh, A. Trend and interannual variability of Walker, monsoon and Hadley circulations defined by velocity potential in the upper troposphere. *Tellus A* **56**, 250–269 (2004).
8. Webster, P. J. et al. Monsoons: Processes, predictability, and the prospects for prediction. *J. Geophys. Res.* **103**(C7), 14451–14510 (1998).
9. Deser, C. & Wallace, J. M. Large-scale atmospheric circulation features of warm and cold episodes in the tropical Pacific. *J. Clim.* **3**, 1254–1281 (1990).
10. Cane, M. A. & Sarachik, E. S. Forced baroclinic ocean motions, Part II: The linear equatorial bounded case. *J. Mar. Res.* **35**, 395–432 (1977).
11. Barber, R. T. & Chavez, F. P. Biological consequences of El Niño. *Science* **222**, 1203–1210 (1983).
12. Trenberth, K. E., Fasullo, J. & Smith, L. Trends and variability in column integrated atmospheric water vapor. *Clim. Dyn.* **24**, doi:10.1007/s00382-005-0017-4 (2005).
13. Soden, B. J., Jackson, D. L., Ramaswamy, V., Schwarzkopf, M. D. & Huang, X. The radiative signature of upper tropospheric moistening. *Science* **310**, 841–844, doi:10.1126/science.1115602 (2005).
14. Boer, G. J. Climate change and the regulation of the surface moisture and energy budgets. *Clim. Dyn.* **8**, 225–239 (1993).
15. Allen, M. R. & Ingram, W. J. Constraints on future changes in the hydrological cycle. *Nature* **419**, 224–228 (2002).
16. Delworth, T. L. et al. GFDL's CM2 global coupled climate models—Part 1: Formulation and simulation characteristics. *J. Clim.* **19**(5), 643–674 (2006).
17. Wittenberg, A. T., Rosati, A., Lau, N.-C. & Ploshay, J. J. GFDL's CM2 global coupled climate models—Part 3: Tropical Pacific climate and ENSO. *J. Clim.* **19**(5), 698–722 (2006).
18. Stouffer, R. et al. GFDL's CM2 global coupled climate models—Part 4: Idealized climate response. *J. Clim.* **19**(5), 723–740 (2006).
19. Clarke, A. J. & Lebedev, A. Long-term changes in equatorial Pacific trade winds. *J. Clim.* **9**, 1020–1029 (1996).
20. Jin, F.-F. An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmos. Sci.* **54**, 811–829 (1997).
21. Fedorov, A. V. & Philander, S. G. Is El Niño changing? *Science* **288**, 1997–2002 (2000).
22. Harrison, D. E. & Vecchi, G. A. El Niño and La Niña—Equatorial Pacific thermocline and sea surface temperature anomalies, 1986–98. *Geophys. Res. Lett.* **28**, 1051–1054 (2001).
23. Cardone, V. J., Greenwood, J. G. & Cane, M. A. On trends in historical marine wind data. *J. Clim.* **3**, 113–127 (1990).
24. Ramage, C. S. Can shipboard measurements reveal secular changes in tropical air-sea heat flux? *J. Clim. Appl. Meteorol.* **23**, 187–193 (1984).
25. Whysall, K. D. B., Cooper, N. S. & Bigg, G. R. Long-term changes in tropical Pacific surface wind field. *Nature* **327**, 216–219 (1987).
26. Harrison, D. E. Post World War II trends in tropical Pacific surface trades. *J. Clim.* **2**, 1561–1563 (1989).
27. Worley, S. J., Woodruff, S. D., Reynolds, R. W., Lubker, S. J. & Lot, N. ICOADS release 2.1 data and products. *Int. J. Climatol.* **25**, 823–842 (2005).
28. Barnett, T. & Parker, D. *Development of the Global Mean Sea Level Pressure Data Set GMSLP2* (Climate Research Technical Note 79, Hadley Centre, Met Office, Exeter, UK, 1997).
29. Kaplan, A., Kushnir, Y. & Cane, M. A. Reduced space optimal interpolation of historical marine sea level pressure. *J. Clim.* **13**, 2987–3002 (2000).
30. Derber, J. & Rosati, A. A global oceanic data assimilation system. *J. Phys. Oceanogr.* **19**, 1333–1347 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G.A.V. was supported by the Visiting Scientist Program at the NOAA/GFDL administered by UCAR. We are grateful to the model development teams at GFDL, and thank A. E. Johansson, M. P. Vecchi, T. Knutson, T. Delworth and J. Russell for comments and suggestions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.A.V. (Gabriel.A.Vecchi@noaa.gov).

A primitive fish provides key characters bearing on deep osteichthyan phylogeny

Min Zhu¹, Xiaobo Yu², Wei Wang¹, Wenjin Zhao¹ & Liantao Jia¹

Osteichthyans, or bony vertebrates, include actinopterygians (teleosts and their relatives) and sarcopterygians (coelacanths, lungfishes and tetrapods). Despite features found in basal actinopterygians (for example, *Dialipina* and *Ligulalepis*)^{1–3} and basal sarcopterygians (for example, *Psarolepis* and *Achoania*)^{4,5}, the morphological gap between the two lineages remains wide and how sarcopterygians developed a dermal surface covering known as cosmine (composed of a pore–canal network and a single layer of odontodes and enamel) is still poorly known^{6–10}. Here we describe a primitive fossil fish, *Meemannia eos* gen. et sp. nov., that possesses an actinopterygian-like skull roof and a cosmine-like dermal surface combining a pore–canal network (found in various fossil sarcopterygians) with superimposed layers of odontodes and enamel (previously known in actinopterygians and some acanthodians^{11–13}). This 405-million-year-old fish from the Lower Devonian of Yunnan (China) demonstrates that cosmine in many fossil sarcopterygians arose step by step through the acquisition of a pore–canal network followed by the subsequently developed ability to resorb previous generations of odontodes and enamel. *Meemannia* provides key characters for studying deep osteichthyan phylogeny and indicates a possible morphotype for the common ancestor of actinopterygians and sarcopterygians.

Sarcopterygii (Romer, 1955)
Meemannia eos gen. et sp. nov.

Diagnosis. A basal osteichthyan fish with a cosmine-like dermal surface composed of a pore–canal network and up to four superimposed layers of odontodes and enamel. Parietals anteriorly separated by an anteriorly broad, triangular postrostral. Parietal almost twice as long as postparietal. Postparietal longer than wide, flanked by large supratemporal with extensive postero–lateral flange. Supraorbital canal passing through the anterior portion of parietal with no connection to the otic canal. Middle and posterior pit-line close to median line. Oto–occipital with an extensive supraotic cavity connected to a large crescent-shaped posterior dorsal fontanelle.
Type species. *Meemannia eos* sp. nov.

Etymology. Generic name after Meemann Chang for her contributions to paleoichthyology. Specific name from Greek *eos*, ‘dawn’, indicating the rudimentary condition of the cosmine-like surface covering.

Holotype. V14536.1, a fairly complete skull roof, Institute of Vertebrate Paleontology and Paleoanthropology (IVPP), Beijing.

Age and locality. Early Devonian (late Lochkovian), Xitun Formation, Qujing, East Yunnan, China.

Description. *Meemannia eos* gen. et sp. nov. (Fig. 1) is represented by three fairly complete skull roofs (V14536.1–3) and one posterior portion of the skull roof (V14536.4) with incompletely preserved

oto–occipital structures. The skull roof surface is punctured by coarse pore openings (though smaller than those in primitive sarcopterygians *Psarolepis*⁴, *Achoania*⁵ and *Styloichthys*¹⁴) often arranged in parallel grooves with intervening smooth ridges.

Meemannia manifests skull roof features variously found in primitive actinopterygians such as *Dialipina*^{1,15}, *Cheirolepis*^{16,17}, *Moythomasia*¹⁸ and *Howqualepis*¹⁹ (Fig. 1a–d, g, h). The parietal is flanked by an elongated dermosphenotic traversed by the anterior portion of the otic canal before it exits at a point close to the posterior corner of the orbital margin. The postparietal is flanked by a large supratemporal with a postero–laterally protruding flange that contributes to a markedly embayed posterior margin of the skull roof. As in *Howqualepis*, the anterior portion of the parietal carries an elongated indentation representing an opening of the supraorbital canal. As in *Dialipina*, the large shield-shaped postrostral anteriorly separates the parietals, the supraorbital canal traverses the anterior portion of the parietal without contacting the otic canal, the middle and posterior pit-lines lie close to the median line, and the rostral part of the skull roof (though unpreserved) seems easily separable from the rest of the skull during preservation, a condition also found in some placoderms (for example, acanthothoracids)^{13,20} and in ‘loose-nosed’ lungfishes²⁰.

Meemannia’s endocranial features revealed in the oto–occipital region (Fig. 1e, f) resemble those in both basal actinopterygians (for example, *Ligulalepis*^{2,3}) and basal sarcopterygians (for example, *Psarolepis*²¹), such as the position and shape of the semicircular canals and utricular recess, a large lozenge-shaped myelencephalic portion of the brain cavity (cav.cr) posteriorly closed off by medially converging ridges, paired lateral cranial canals (lcc), a large supraotic cavity (cav.so) with anteriorly diverging extensions, and a large posterior dorsal fontanelle (pdf). The last two may be general gnathostome features because they also exist in some chondrichthyans²².

The most remarkable feature of *Meemannia* is a unique histological condition bearing on the origin of cosmine, an enigmatic vertebrate hard tissue unknown in living forms but widely spread among fossil sarcopterygians^{13,23}. As revealed by 15 transverse sections of specimen V14534.3 (Fig. 2a–c), the upper portion consists of three or four superimposed layers of enamel (e1–e4) and odontodes (od1–od4) separated by flask-shaped pore cavities (pc), interconnecting horizontal canals (hc), and pore openings (p). The middle portion consists of vascular bone and is poorly developed, whereas the lower portion consisting of lamellar bone is well developed and often directly underlies the upper portion.

The enamel of the most superficial layer dips slightly into the pore openings, whereas the enamel of each underlying layer dips slightly into the side wall of the pore cavities at different levels. This arrangement, together with the superimposition of three or four enamel–odontode layers, indicates that each layer (or successive generation)

¹Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. ²Department of Biological Sciences, Kean University, Union, New Jersey 07083, USA.

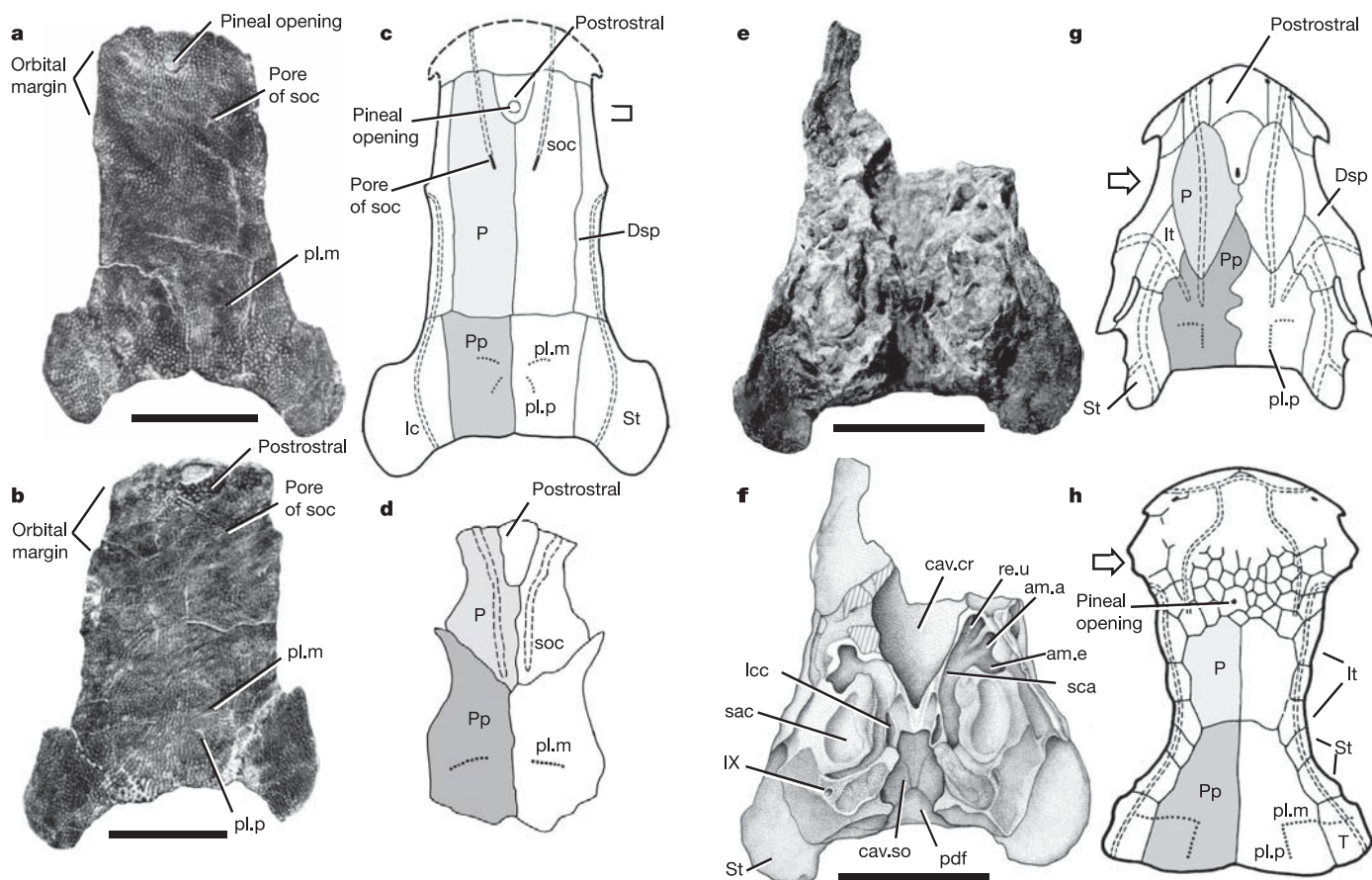


Figure 1 | *Meemannia eos* gen. et sp. nov. This 405-Myr-old fish shows a mixture of basal actinopterygian and sarcopterygian features. **a, b**, Dorsal view of skull roof (**a**, holotype, V14536.1; **b**, V14536.2). **e, f**, Ventral view of posterior portion of the skull roof with incompletely preserved oto-occipital structures (**e**, V14536.4; **f**, illustrative drawing). **c, d, g, h**, Reconstruction of skull roof (**c**) compared with two actinopterygians, *Dialipina*¹⁵ (**d**) and *Cheirolepis*¹⁶ (**g**), and one sarcopterygian, *Powichthys*³⁰ (**h**). Abbreviations: am.a, am.e, anterior and external ampullae; cav.cr, cranial cavity; cav.so,

supraotic cavity; Dsp, dermosphenotic; It, intertemporal; lc, otic portion of the main lateral line canal; lcc, lateral cranial canal; P, parietal; pdf, posterior dorsal fontanelle; pl.m, pl.p, middle and posterior pit-lines; Pp, postparietal; re.u, utricular recess; sac, sacculus; sca, anterior semicircular canal; soc, supraorbital canal; St, supratemporal; T, tabular; IX, exit of the ninth cranial nerve. Open arrow in **c, g** and **h** indicates the position of the orbit. Scale bar, 5 mm (**a, b, e, f**).

might be deposited without resorbing previously deposited layers and that the pore cavities grow in depth with the successive deposition of each layer.

The pore cavities, horizontal canals and pore openings in *Meemannia* represent typical components of a pore–canal network characteristic of cosmine found in many crown-group sarcopterygians. However, the superimposition of enamel–odontode layers as well as the weak development of the vascular bone resembles the dermal bone features found in actinopterygians (for example, *Andreolepis*¹¹) and some acanthodians¹². Cosmine in crown-group sarcopterygians has one single layer of odontodes and enamel, and a process of resorption preceding each event of cosmine redeposition prevents the formation of overlapping layers⁷. Although isolated cases of buried odontodes have been reported in *Porolepis*²⁴ (a porolepiform) and *Uranolophus*²⁵ (a lungfish), these occur below the level of the cosmine layer and no superimposed layers are formed. Continuing histological work reveals that *Psarolepis* and *Styloichthys* also possess a pore–canal network embedded in superimposed layers of odontodes and enamel (Fig. 2d, e). However, the superimposed layers occur less frequently than in *Meemannia* and the number of superimposed layers varies in neighbouring regions, indicating the possible initial development of partial resorption in these two forms.

To explore the phylogenetic position of *Meemannia*, we constructed a data matrix of 125 characters (with 25 available characters for *Meemannia*) and 19 taxa representing both actinopterygians

and sarcopterygians. Phylogenetic analysis²⁶ (see Supplementary Information) yields two most parsimonious trees, both showing *Meemannia* as the most basal sarcopterygian below the node of *Psarolepis* (Fig. 3). The two trees differ only in the position of *Ligulalepis*, which forms the sister-group of either *Dialipina* or *Osorioichthys*²⁷. However, this result should be treated with caution because the node below *Meemannia* and the nodes in the actinopterygian lineage have low Bremer support values.

Despite the tentative nature of *Meemannia*'s position, its unique character combination has wide implications for studying deep osteichthyan phylogeny and the early divergence of actinopterygians and sarcopterygians.

First, *Meemannia* makes it possible to interpret features found in basal members of both lineages in the common framework of deep osteichthyan phylogeny. For instance, the apparent separation of the rostral part of the skull roof from the rest of the skull (in *Dialipina* and *Meemannia*), the *Psarolepis*-like endocranial features (in *Ligulalepis*^{2,3} and *Meemannia*), the actinopterygian-like skull roof pattern (in *Meemannia*, *Dialipina* and other actinopterygians), the absence of a dermal joint between the parietal and the postparietal (in *Meemannia*, *Dialipina* and other actinopterygians), and the superimposition of enamel–odontode layers (in *Meemannia* and *Andreolepis*¹¹, as well as in *Psarolepis* and *Styloichthys*) should now be viewed as general osteichthyan features existing in the common ancestor of actinopterygians and sarcopterygians.

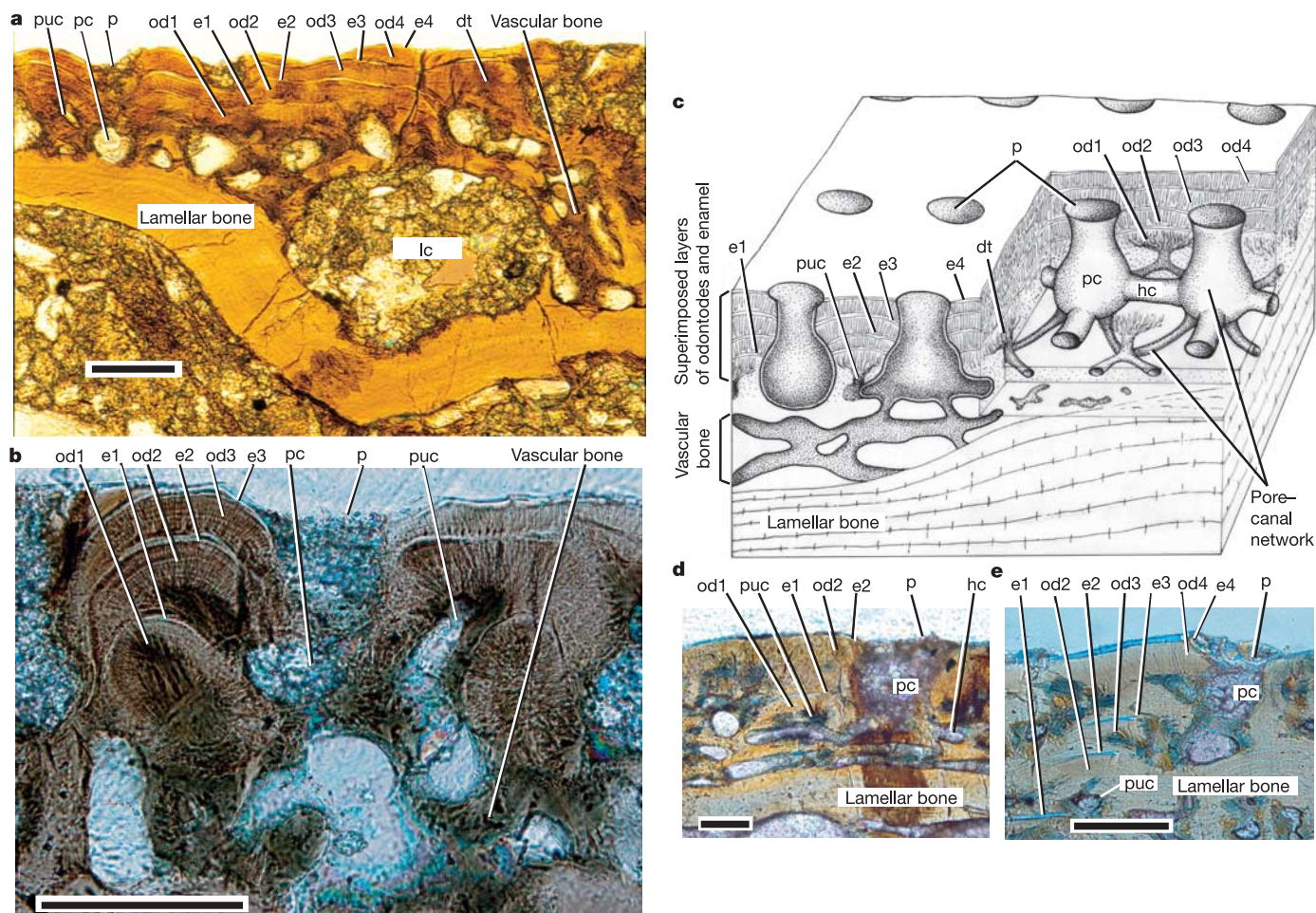


Figure 2 | *Meemannia* possesses a pore-canal network combined with three or four layers of odontodes and enamel in dermal bone surface. This condition preceded the appearance of typical cosmine found in crown-group sarcopterygians. *Psarolepis* and *Styloichthys* present an intermediate condition between *Meemannia* and crown-group sarcopterygians. **a**, **b**, Two transverse sections of *Meemannia* skull roof (V14534.3). **c**, Reconstruction

based on transverse sections of V14534.3. **d**, Transverse section of *Psarolepis* (V14600.5, skull roof). **e**, Transverse section of *Styloichthys* (V14599.4, cleithrum). Abbreviations: dt, dentine tubules; e1–e4, layers of enamel; hc, horizontal canal; lc, otic portion of the main lateral line canal; od1–od4, layers of odontodes; p, pore opening; pc, pore cavity; puc, pulp cavity. Scale bar, 100 μ m.

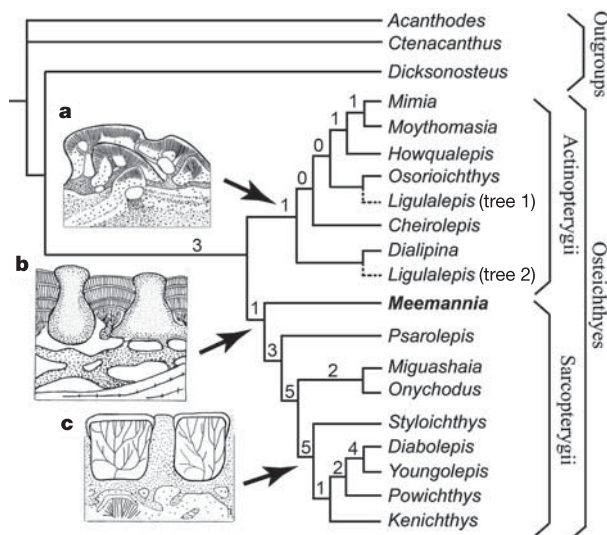


Figure 3 | Cladogram showing the tentative position of *Meemannia* as the most basal sarcopterygian and the step-by-step origin of cosmine in crown-group sarcopterygians. The figure is based on two most parsimonious trees that differ in the positions of *Ligulalepis*. Bremer support values are shown at nodes. Tree length 222, consistency index

0.6216, homoplasy index 0.3784, retention index 0.7807, rescaled consistency index 0.4853. See Supplementary Information for details. Insets compare the histological features of *Meemannia* (**b**) with those found in actinopterygians (*Andreolepis*¹¹, **a**) and crown-group sarcopterygians (*Porolepis*²⁴, **c**).

Second, *Meemannia* indicates that cosmine in crown-group sarcopterygians might have developed step by step, first through the acquisition of a pore–canal network (in *Meemannia*) and subsequently through the development of increased ability to resorb previously deposited enamel–odontode layers (in *Psarolepis*, *Styloichthys* and crown-group sarcopterygians). *Psarolepis* and *Styloichthys* may represent an intermediate stage in which the ability to resorb previous layers was initially developed before it became fully established in crown-group sarcopterygians.

Last, *Meemannia* may shed light on long-standing controversies about the biology of cosmine^{6,7}. With three or four superimposed layers of odontodes and enamel associated with only one pore–canal network, *Meemannia* shows that there is no one-to-one association between the deposition of each layer of odontodes and enamel and the formation of a pore–canal system. On the contrary, the slight dipping of successive enamel layers into the side wall of pore cavities indicates that the space known as pore cavities increases in depth as each enamel–odontode layer is deposited. Although opinions still differ on the biological interpretation of the pore–canal network^{6,7,28,29}, *Meemannia* seems to lend support to the theory that the pore–canal network might represent vascular structures involved in the deposition of odontodes and enamel¹⁰, rather than a network of cutaneous sensory or glandular structures^{7,9,20}.

METHODS

Phylogenetic analysis was performed with the phylogenetic package PAUP*4.0b10 (ref. 26). See Supplementary Information for the list of 125 characters with sources of reference, the data matrix, the two most parsimonious trees and the characters defining major clades. As the position of *Meemannia* is based on only 25 available characters, the phylogenetic result is subject to future changes when more characters become available.

Received 26 July 2005; accepted 4 January 2006.

- Schultze, H.-P. & Cumbaa, S. L. in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny and Development* (ed. Ahlberg, P. E.) 315–332 (Taylor & Francis, London, 2001).
- Basden, A. M., Young, G. C., Coates, M. I. & Ritchie, A. The most primitive osteichthyan braincase? *Nature* **403**, 185–188 (2000).
- Basden, A. M. & Young, G. C. A primitive actinopterygian neurocranium from the Early Devonian of southeastern Australia. *J. Vertebr. Paleontol.* **21**, 754–766 (2001).
- Zhu, M., Yu, X.-B. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397**, 607–610 (1999).
- Zhu, M., Yu, X.-B. & Ahlberg, P. E. A primitive sarcopterygian fish with an eyestalk. *Nature* **410**, 81–84 (2001).
- Ørvig, T. Cosmine and cosmine growth. *Lethaia* **2**, 241–260 (1969).
- Thomson, K. S. On the biology of cosmine. *Yale Univ. Peabody Mus. Nat. Hist. Bull.* **40**, 1–59 (1975).
- Rosen, D. E., Forey, P. L., Gardiner, B. G. & Patterson, C. Lungfishes, tetrapods, paleontology, and plesiomorphy. *Bull. Am. Mus. Nat. Hist.* **167**, 159–276 (1981).
- Meinke, D. K. A review of cosmine: its structure, development, and relationship to other forms of the dermal skeleton in osteichthyans. *J. Vertebr. Paleontol.* **4**, 457–470 (1984).
- Bemis, W. E. & Northcutt, R. G. Skin and blood vessels of the snout of the Australian lungfish, *Neoceratodus forsteri*, and their significance for interpreting the cosmine of Devonian lungfishes. *Acta Zool. (Stockh.)* **73**, 115–139 (1992).
- Gross, W. Fragliche Actinopterygier-Schuppen aus dem Silur Gotlands. *Lethaia* **1**, 184–218 (1968).
- Gross, W. Downtonische und dittonische Acanthodier-Reste des Ostseegebietes. *Palaeontogr. Abt. A* **136**, 1–82 (1971).
- Janvier, P. *Early Vertebrates* (Clarendon, Oxford, 1996).
- Zhu, M. & Yu, X.-B. A primitive fish close to the common ancestor of tetrapods and lungfish. *Nature* **418**, 767–770 (2002).
- Schultze, H.-P. in *Fossil Fishes as Living Animals* (ed. Mark-Kurik, E.) 233–242 (Academy of Sciences of Estonia, Tallinn, 1992).
- Pearson, D. M. & Westoll, T. S. The Devonian actinopterygian *Cheirolepis* Agassiz. *Trans. R. Soc. Edinb.* **70**, 337–399 (1979).
- Arratia, G. & Cloutier, R. in *Devonian Fishes and Plants of Miguasha, Quebec, Canada* (eds Schultze, H.-P. & Cloutier, R.) 165–171 (Dr Friedrich Pfeil, München, 1996).
- Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of Western Australia. *Bull. Br. Mus. Nat. Hist. (Geol.)* **37**, 173–428 (1984).
- Long, J. A. New palaeoniscoid fishes from the Late Devonian and Early Carboniferous of Victoria. *Mem. Assoc. Australas. Palaeontol.* **7**, 1–64 (1988).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic, London, 1980).
- Yu, X.-B. A new porolepiform-like fish *Psarolepis romeri* gen. et sp. nov. (Sarcopterygii, Osteichthyes) from the Lower Devonian of Yunnan, China. *J. Vertebr. Paleontol.* **18**, 261–274 (1998).
- Maisey, J. G. in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, Genetics and Development* (ed. Ahlberg, P. E.) 263–288 (Taylor & Francis, London, 2001).
- Miles, R. S. Dipnoan (lungfish) skulls and the relationships of the group: a study based on new species from the Devonian of Australia. *Zool. J. Linn. Soc.* **61**, 1–328 (1977).
- Gross, W. Über Crossopterygier und Dipnoer aus dem baltischen Oberdevon im Zusammenhang einer vergleichenden Untersuchung des Porenkanalsystems paläozoischer Agnathen und Fische. *K. Svensk. Vet. Akad. Handl.* **5**(4), 1–140 (1956).
- Meinke, D. K. in *The Biology and Evolution of Lungfishes* (eds Bemis, W. E., Burggren, W. W. & Kemp, N. E.) 133–149 (Alan R. Liss, New York, 1987).
- Swofford, D. L. *PAUP*: Phylogenetic analysis using parsimony (* and other methods)*, version 4.0b10 (Sinauer Associates, Sunderland, Massachusetts, 2003).
- Taverne, L. *Osorioichthys marginis*, 'Paléonisciforme' du Famennien de Belgique et la phylogénie des Actinoptérygiens dévoniens (Pisces). *Bull. Inst. R. Sci. Nat. Belg. Sci. Terre* **67**, 57–78 (1997).
- Smith, M. M. & Hall, B. K. Development and evolutionary origins of vertebrate skeletogenic and odontogenic tissues. *Biol. Rev.* **65**, 277–373 (1990).
- Thomson, K. S. in *Problems in Vertebrate Evolution* (eds Andrews, S. M., Miles, R. S. & Walker, A. D.) 247–271 (Academic, London, 1977).
- Jessen, H. L. A new choanate fish, *Powichthys thorsteinssoni* n. g., n. sp., from the Early Lower Devonian of the Canadian Arctic Archipelago. *Coll. Cent. Natl. Rech. Scient.* **218**, 213–222 (1975).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. E. Ahlberg for advice and useful discussions, M. Yang for artwork, J. Zhang for photographic work, X. Lu for specimen preparation, and W. Zhang for making thin sections. This work was supported by the Chinese Foundation of Natural Sciences, the Special Funds for Major State Basic Research Projects of China, and the Chinese Academy of Sciences. X.Y. thanks Kean University for faculty research and development support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Z. (zhumin@ivpp.ac.cn).

Climate change and population declines in a long-distance migratory bird

Christiaan Both^{1,2}, Sandra Bouwhuis^{1†}, C. M. Lessells¹ & Marcel E. Visser¹

Phenological responses to climate change differ across trophic levels^{1–3}, which may lead to birds failing to breed at the time of maximal food abundance. Here we investigate the population consequences of such mistiming in the migratory pied flycatcher, *Ficedula hypoleuca*⁴. In a comparison of nine Dutch populations, we find that populations have declined by about 90% over the past two decades in areas where the food for provisioning nestlings peaks early in the season and the birds are currently mistimed. In areas with a late food peak, early-breeding birds still breed at the right time, and there is, at most, a weak population decline. If food phenology advances further, we also predict population declines in areas with a late food peak, as in these areas adjustment to an advanced food peak is insufficient⁴. Mistiming as a result of climate change is probably a widespread phenomenon¹, and here we provide evidence that it can lead to population declines.

Ongoing climate change leaves a clear global fingerprint on ecosystems. Many organisms bring forward the timing of their seasonal activities, whether it be flowering in plants, budding of trees, emergence of insects or breeding in birds^{5–7}. Despite this general advancement, some species may not cope with climate change because their response differs from the response of organisms at lower levels of the food chain^{1–4,8,9}, leading to a mismatch between the timing of reproduction and the main food supply¹⁰. This mistiming can have a clear effect on species population dynamics and ecosystem functioning^{2,11}. In general, we expect the populations that are most mistimed to decline most in number. Here we show how populations of a small passerine bird have declined as a consequence of climate change, because the phenology of their main food supply during breeding has advanced more than the birds' breeding date.

We studied the population ecology of the long-distance migratory passerine, the pied flycatcher *Ficedula hypoleuca*, and its caterpillar food supply. We have previously shown in this long-term study in the Netherlands that the flycatchers have advanced their laying date but not the timing of their spring arrival in the Netherlands, and that the advancement in laying date was not sufficient to track the advancement of spring, leading to increased selection for early breeding⁴. The temperate forest habitat of our study area is characterized by a clear peak in caterpillar abundance in spring, and caterpillars are an important food source for nestling flycatchers^{12,13}. The timing of this caterpillar peak differs between areas (see Supplementary Information) and years, with a clear shift forward over the past 20 years in our main study population¹⁴.

We predicted that areas with increased mismatch between the timing of the birds and the peak availability of their prey would show a strong population decline. To test this prediction, we collated annual population counts between 1987 and 2003 from ten nest box populations in the Netherlands that differed strongly in population

trends. If increased mistiming is the cause of population declines, we predict that populations in the areas with the earliest food peak will have declined most strongly. This is because these long-distance migrants have a relatively fixed spring migration programme¹⁵, and in early food phenology areas these birds have a shorter period between their arrival and the time of the food peak. A short time interval between arrival and breeding may act as a constraint, because the birds can not shorten this much further. We expect that populations in these areas of early food peak might also react less flexibly to increases in temperature, and consequently decline in number.

We found strong support for our hypothesis: pied flycatchers have declined by about 90% in areas with the earliest food peaks, but have only declined by about 10% in areas with the latest food peaks (Fig. 1a). Because we measured the caterpillar peak date in only one year (2003, at the end of the study period), we used the percentage of great tits producing a second brood over a six-year period (1985–1990) as a second measure of the phenological state of the area, because great tits produce second broods only if caterpillar peaks are late¹⁶. Flycatcher populations declined most in areas with

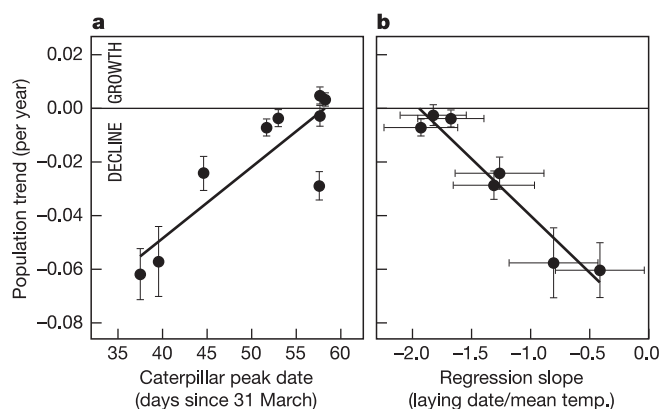


Figure 1 | Population trends of Dutch pied flycatcher populations. **a**, **b**, Trends in response to the local date of the caterpillar peak (in days since 31 March) (Spearman rank correlation: $r_s = 0.80$, $n = 9$, $P = 0.013$) (**a**), and the slope of annual median laying date on spring (16 April–15 May) temperature ($r_s = -0.86$, $n = 7$, $P = 0.03$) (**b**). Populations of pied flycatchers with an early food peak and a weak response declined most strongly. Population trend is the slope of the regression of the log number of breeding pairs against year. In **b**, the x axis shows the slope of a linear regression of median laying date against mean temperature from 16 April–15 May. Error bars represent the standard errors of the slopes of the regression lines. All points in **b** are also in **a**, except for one point, for which we had no data regarding the caterpillar peak.

¹Netherlands Institute of Ecology (NIOO-KNAW), PO Box 40, 6666ZG Heteren, The Netherlands. ²Animal Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 14, 9750AA Haren, The Netherlands. [†]Present addresses: Behavioural Biology Group, University of Groningen, PO Box 14, 9750AA Haren, The Netherlands, and The Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

the lowest percentage of great tit second broods (Spearman rank correlation: $r_s = 0.94$, $n = 6$ populations, $P = 0.003$), confirming our result based on direct measures of caterpillar peak dates.

If migration timing does act as a constraint, we expect that the flycatcher breeding would be least able to advance with increasing temperatures in areas with the earliest food phenology. Indeed, the populations that adjusted their laying date least to temperature were the ones with the strongest decline in population numbers (Fig. 1b). This is probably not due to genetic differences in reaction norms between populations because such differences were not found in a related species¹⁷, and the exchange of ringed birds occurs between some of our study populations. The population declines in areas with early food peaks and in populations least flexible to increasing temperatures thus strongly support the idea that these declines are attributable to insufficient adjustment of arrival and laying dates to climate change.

The decline in the number of pied flycatcher breeding pairs in areas with an early food phenology was not due to general deterioration of the habitat, because caterpillar biomass was highest in areas with an early food peak (Spearman rank correlation between caterpillar peak abundance and peak date: $r_s = -0.695$, $n = 9$, $P < 0.05$), and population trends of resident great tits (*Parus major*) breeding in the same nest boxes were unrelated to the date of the caterpillar peak (linear regression: $F_{1,6} = 0.38$, $P = 0.56$; slopes differed between species: $F_{1,13} = 11.66$, $P = 0.013$), despite their dependence on caterpillars for feeding chicks¹⁸.

We have thus clearly shown that climate-change-induced mistiming leads to population declines in a migratory songbird. One possible reason why pied flycatchers have not adjusted sufficiently to climate change is that their arrival from their wintering grounds has not advanced and probably acts as a constraint on laying dates^{4,19}. Because we have also shown this mistiming in an area with a late food phenology^{4,20}, we expect flycatcher populations in these areas to decline if the food peak advances further. The decline in areas with an early food phenology is probably not just the result of changed habitat selection, because in two-thirds of the Netherlands declines of about 50% were reported between 1986 and 1999 in the Dutch common bird census²¹, which is in agreement with the general decline in European long-distance migrants²². Furthermore, since 1988 declines have been reported in 45 out of 62 nest box areas in the UK (J. Wright and the British Trust for Ornithology, personal communication), suggesting that these declines are more widespread, but differ across sites. In general, we expect climate change to be a greater threat to long-distance migrants than to resident species, and to species breeding in strongly seasonal environments than to species living in less seasonal habitats²³. Apart from a mismatch due to an advance in the timing of their food, the duration of high food availability may decrease because caterpillars grow faster and advance pupation at higher temperatures²⁴. The general decline in many long-distance migratory species in both Europe²² and North America²⁵ may thus be particularly pronounced in seasonal habitats and may be exacerbated by climate change, which can exaggerate the temporal mismatch between avian predators and their prey.

METHODS

We studied ten pied flycatcher populations in the Netherlands separated by 3–150 km (see Supplementary Information). Population size (the number of pairs using the nest boxes) was analysed from 1987 (the first year for which we have data for all populations) until 2003. The number of pairs using the nest boxes is used here as a population size for the area, because in the Netherlands about 90% of the species breeds in nest boxes, and in nest box areas more than 98% of flycatcher pairs breed in the boxes (unpublished observation). The population trend is the slope of the regression with $\log_{10}(\text{population size} + 1)$ the dependent variable and year as the independent variable, and is a measure of the relative growth or decline of the population.

Laying date was known for six populations, and we calculated the annual median laying date from 1980–2002. Previous work has shown that laying date correlates strongly with spring temperature, and we used the slope of the linear

regression of the annual median laying date (dependent variable) and mean temperature from 16 April–15 May (independent variable) as measure of the plasticity of laying date⁴. The median laying date in the early part of the study is 15 May, and we showed that laying date is strongly correlated with temperature over this period in 25 study areas across Europe²⁶. Areas differed in the period for which laying dates were available; we used all the available data in order to get the most accurate measure of plasticity. We calculated median laying dates only for years in which more than seven pairs bred in each population.

In 2003, we measured the caterpillar peak for two pedunculate oak *Quercus robur* trees in each of nine areas. Representative trees were selected in the main breeding areas of the pied flycatchers. We placed a 50 cm × 50 cm cloth net under each tree to catch caterpillar droppings, and collected the samples in the nets every five days. Caterpillar biomass in the trees was estimated from the dry weight of the droppings²⁷ and the caterpillar peak date was defined as the middle day in the five-day period for which caterpillar biomass was maximal. In the Netherlands, 2003 was a warm spring, and the caterpillar peak in our main study area was the third-earliest since we started measuring it in 1985 (ref. 14). Using a regression slope, the peak advanced 16 days from 1985 to 2003.

As a second approximation of the general phenological state of each area, we used the percentage of great tits producing second broods. This species is known to produce more second broods if the caterpillar peak is late¹⁶, and the percentage of second broods is thus a measure of whether caterpillar peaks are early or late. We took the average of the annual proportions for the years 1985–1990, which was at the start of our flycatcher population analysis, because as a result of climate change the proportion of great tits producing second broods has declined in some habitats²⁸. The proportion of second broods was known for six of the ten study populations.

Received 17 October; accepted 22 December 2005.

- Visser, M. E. & Both, C. Shifts in phenology due to global climate change: the need for a yardstick. *Proc. R. Soc. B* **272**, 2561–2569 (2005).
- Stenseth, N. C. *et al.* Ecological effects of climate fluctuations. *Science* **297**, 1292–1296 (2002).
- Merila, J., Kruuk, L. E. B. & Sheldon, B. C. Cryptic evolution in a wild bird population. *Nature* **412**, 76–79 (2001).
- Both, C. & Visser, M. E. Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. *Nature* **411**, 296–298 (2001).
- Walther, G. R. *et al.* Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002).
- Parnesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
- Root, T. L. *et al.* Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60 (2003).
- Visser, M. E., van Noordwijk, A. J., Tinbergen, J. M. & Lessells, C. M. Warmer springs lead to mistimed reproduction in great tits (*Parus major*). *Proc. R. Soc. Lond. B* **265**, 1867–1870 (1998).
- Visser, M. E. & Holleman, L. J. M. Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proc. R. Soc. Lond. B* **268**, 289–294 (2001).
- Visser, M. E., Both, C. & Lambrechts, M. M. Global climate change leads to mistimed avian reproduction. *Adv. Ecol. Res.* **35**, 89–110 (2004).
- McLaughlin, J. F., Hellmann, J. J., Boggs, C. L. & Ehrlich, P. R. Climate change hastens population extinctions. *Proc. Natl Acad. Sci. USA* **99**, 6070–6074 (2002).
- von Haartman, L. Der Trauerfliegenschneider. 3. Die Nahrungsbiologie. *Ann. Zool. Fenn.* **83**, 1–95 (1954).
- Sanz, J. J. Effect of habitat and latitude on nestling diet of Pied Flycatchers *Ficedula hypoleuca*. *Ardea* **86**, 81–88 (1998).
- Visser, M. E., Holleman, L. J. M. & Gienapp, P. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia* **147**, 164–172 (2006).
- Gwinner, E. Circannual clocks in avian reproduction and migration. *Ibis* **138**, 47–63 (1996).
- Verboven, N., Tinbergen, J. M. & Verhulst, S. Food, reproductive success and multiple breeding in the Great Tit, *Parus major*. *Ardea* **89**, 387–406 (2001).
- Brommer, J. E., Merila, J., Sheldon, B. C. & Gustafsson, L. Natural selection and genetic variation for reproductive reaction norms in a wild bird population. *Evolution Int. J. Org. Evolution* **59**, 1362–1371 (2005).
- van Balen, J. H. A comparative study of the breeding ecology of the Great Tit *Parus major* in different habitats. *Ardea* **61**, 1–93 (1973).
- Both, C., Bijlsma, R. G. & Visser, M. E. Climatic effects on timing of spring migration and breeding in a long-distance migrant, the pied flycatcher *Ficedula hypoleuca*. *J. Avian Biol.* **36**, 368–373 (2005).
- Both, C. & Visser, M. E. The effect of climate change on the correlation between avian life-history traits. *Glob. Change Biol.* **11**, 1606–1613 (2005).
- Boele, A., van Dijk, A. J., Hustings, F. & Zoetebier, D. Regional differences in the development of European Pied Flycatcher *Ficedula hypoleuca* numbers. *Limosa* **74**, 103–115 (2001).

22. Berthold, P., Fiedler, W., Schlenker, R. & Querner, U. 25-year study of the population development of Central European songbirds: a general decline, most evident in long-distance migrants. *Naturwissenschaften* **85**, 350–353 (1998).
23. Winkler, D. W., Dunn, P. O. & McCulloch, C. E. Predicting the effects of climate change on avian life-histories. *Proc. Natl Acad. Sci. USA* **99**, 13595–13599 (2002).
24. Buse, A., Dury, S. J., Woodburn, R. J. W., Perrins, C. M. & Good, J. E. G. Effects of elevated temperature on multi-species interactions: the case of Pedunculate Oak, Winter Moth and Tits. *Funct. Ecol.* **13** (suppl.), 74–82 (1999).
25. Sauer, J. R., Pendleton, G. W. & Peterjohn, B. G. Evaluating causes of population change in North American insectivorous songbirds. *Cons. Biol.* **10**, 465–478 (1996).
26. Both, C. *et al.* Large-scale geographical variation confirms that climate change causes birds to lay earlier. *Proc. R. Soc. Lond. B* **271**, 1657–1662 (2004).
27. Tinbergen, J. M. & Dietz, M. W. Parental energy expenditure during brood rearing in the Great Tit (*Parus major*) in relation to body mass, temperature, food availability and clutch size. *Funct. Ecol.* **8**, 563–572 (1994).
28. Visser, M. E. *et al.* Variable responses to large-scale climate change in European *Parus* populations. *Proc. R. Soc. Lond. B* **270**, 367–372 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The long-term data on the Hoge Veluwe were collected under the directorships of H. van Balen and A. van Noordwijk, and the database was managed by J. Visser. B. van der Brink, B. Blaauw, H. and A. Dekhuijzen, H. Jansen, and J. van Laar provided population data. We are grateful to Nationaal Park de Hoge Veluwe, Staatsbosbeheer, Natuurmonumenten and Het Geldersch Landschap for permission to work on their properties. Comments made by J. Harvey, T. Piersma, D. Winkler and J. Wright improved the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.B. (C.Both@rug.nl).

LETTERS

Testing the genetics underlying the co-evolution of mate choice and ornament in the wild

Anna Qvarnström^{1*}, Jon E. Brommer^{2*} & Lars Gustafsson¹

One of the most debated questions in evolutionary biology is whether female choice of males with exaggerated sexual displays can evolve as a correlated response to selection acting on genes coding for male attractiveness or high overall viability. To date, empirical studies have provided support for parts of this scenario¹, but evidence for all key genetic components in a natural population is lacking. Here we use animal-model quantitative genetic analysis^{2,3} on data from over 8,500 collared flycatchers (*Ficedula albicollis*) followed for 24 years to quantify all of the key genetic requirements of both fisherian^{4,5} and 'good-genes' models^{6,7} on sexual selection in the wild. We found significant additive genetic variances of all the main components: male ornament (forehead patch size), female mate choice for this ornament, male fitness and female fitness. However, when the necessary genetic correlations between these components were taken into account, the estimated strength of indirect sexual selection on female mate choice was negligible. Our results show that the combined effect of environmental influences on several components reduces the potential for indirect sexual selection in the wild. This study provides insight into the field of sexual selection by showing that genes coding for mate choice for an ornament probably evolve by their own pathways instead of 'hitchhiking' with genes coding for the ornament.

Although it is well established that mate choice often promotes the evolution of exaggerated display traits (ornaments)⁸, it remains a challenge to explain how mate choice itself evolves¹. According to a group of influential models on sexual selection, males carry genes coding for sexual attractiveness (fisherian models)^{4,5} or viability (good-genes models)^{6,7} that enhance the fitness of their offspring in the population. Both fisherian and good-genes processes assume heritable variation in mate choice, and that females with strong preference mate with highly ornamented males, thereby creating a genetic correlation between mate choice and ornament (and potential good genes linked with the ornament) in the offspring generation. Consequently, mate choice will become a target of indirect selection, and will increase in every generation as a correlated response to selection favouring genes coding for attractiveness or viability.

Empirical understanding of the genetics underlying the co-evolution of mate choice and ornaments remains virtually unexplored¹. Laboratory studies have provided estimates of some components: ornaments and female mate choice have comparatively high heritabilities and are genetically correlated^{9,10}. Artificial selection on male ornaments results in evolution of mate choice^{11,12}, and one study found that mate choice was associated with genes that increased the average weight of offspring¹³. Although these laboratory studies have increased our general knowledge of the possibility of indirect selection on mate choice, there are at least two reasons why a complete test of the theory requires investigation in a natural

population. First, in order to understand the importance of indirect selection in the evolution of mate choice, the strength of indirect selection needs to be assessed, and estimates of fitness (and thus selection) inevitably need to be derived from natural populations¹⁴. Second, genetic variances and covariances depend on the environment in which they are measured¹⁵, which means that estimates obtained in the laboratory may differ from those obtained under natural conditions.

There are four critical parameters in models of indirect sexual selection on female mate choice (see Supplementary Methods). First, female choice (C) and the male ornament (O) should be heritable. Second, there must be additive genetic variance in fitness (W) in the population. Third, female choice must be genetically correlated with the ornament (r_{CO}), and finally, the male ornament must be correlated with fitness (r_{OW}). Here we quantify all of these genetic components of indirect sexual selection in a wild population of collared flycatchers (*Ficedula albicollis*), small migratory birds that breed in central and eastern Europe. As for most birds, males compete over resources necessary for attracting females (in this case, nest sites) and both parents participate in caring for their young. Previous research has documented the importance of the male flycatcher's ornament (a conspicuous white forehead patch) in sexual selection through male–male competition^{16,17}, sperm competition¹⁸ and mate choice¹⁹. Sons inherit their fathers' forehead patch size^{20,21}. Females prefer to mate with males with a large forehead patch, especially late in the season, and receive fitness benefits from doing so¹⁹. Forehead patch size has also been directly implicated in a good-genes process because extra-pair offspring that are sired by highly ornamented males fledge in relatively better condition¹⁸ and therefore have higher survival chances than within-pair offspring. Hence, the evolution of female mate choice for this ornament may be at least partly driven by indirect selection through either a fisherian or good-genes process.

We used an animal model^{2,3} to calculate additive genetic and environmental (co)variances in fitness expressed in both sexes, male forehead patch size, and in female mate choice for this ornament, using long-term data collected on the Swedish island of Gotland over 24 years (see Supplementary Information). As mate choices can only be measured when expressed¹, our approach was to consider the observed size of her partner's forehead patch as an estimate of a female's (phenotypic) mate choice.

In accordance with previous results^{20,21}, we found that the size of a male's forehead patch was an informative signal. The ornament had a relatively strong genetic component (heritability $h^2 = 0.381 \pm 0.028$ ($\pm$ s.e.m.), $P < 0.001$; Fig. 1). We also found evidence for significant additive genetic variance in male and female annual fitness (h^2 of annual fitness, male $h^2 = 0.031 \pm 0.012$, $P = 0.005$; female $h^2 = 0.044 \pm 0.013$, $P = 0.0007$). This estimate of breeding value

¹Animal Ecology, Department of Ecology and Evolution, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. ²Bird Ecology Unit, Department of Biological and Environmental Sciences, University of Helsinki, PO Box 65 (Viksbågen 1), FI-00014 Helsinki, Finland.

*These authors contributed equally to this work.

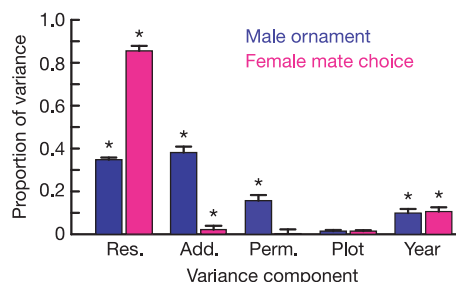


Figure 1 | Components of variance in male ornament and female choice for the ornament. A total of 4,220 male (ornament, blue) and 4,345 female (choice for the ornament, pink) collared flycatchers were measured. The proportion of the total variance in each trait attributed to residual (Res), additive genetic (Add), permanent environment (Perm), study plot and year were calculated using an animal model^{2,3}. Significant components are indicated with an asterisk; error bars denote 1 s.e.m. The male ornament is informative because it is heritable and also signals other individual-specific differences across males. Females show a significant (but weak) genetically coded mate choice with respect to the male's ornament.

for fitness incorporates both adult survival and reproduction. The overall coefficient of variation is $CV_W = 1/2(CV_W^m + r_W^{mf} CV_W^f) = 0.113$ (Table 1), which averages how fitness genes are expressed in males (m) and females (f) by weighting the contribution of the latter by the genetic correlation in fitness between the sexes r_W^{mf} (see Supplementary Information). The overall coefficient of genetic variation in fitness determines the maximum potential change in mate choice through indirect selection, and is specific for the environment in which it is measured. There was a positive but weak genetic correlation between ornament and male annual fitness, $r_{OW} = 0.154 \pm 0.094$ ($P = 0.1$). We do not know of any other study that has calculated this correlation, but in general, such genetic correlations are expected to be low¹⁴. By choosing a mate on the basis of their forehead patch size, females experience positive (although weak) potential indirect selective benefit (of $r_{OW} CV_W = 0.017$ standard deviations per year). However, from the female's perspective, indirect sexual selection has little scope to operate because female mate choice largely lacks an additive genetic basis: females showed very low consistency in their mate choice across individuals and had a significant (but very weak) genetic component of mate choice ($h^2 = 0.026 \pm 0.010$, $P = 0.009$; Fig. 1). The low heritability of mate choice was due to low additive genetic variance (about 20 times lower than the additive genetic variance for the male ornament, see Table 1). The genetic correlation between mate choice and male ornament was close to zero ($r_{CO} = -0.015 \pm 0.169$, $P = 0.93$). All in all, the change in female mate choice attributable to indirect sexual selection is a function of how well genes for mate choice can associate themselves with genes for fitness through the genes for male ornamentation, which is given by the product of all the genetic parameters we estimated here¹⁴. Hence, the expected change in

female mate choice equals $h_{CO} r_{OW} [1/2(CV_W^m + r_W^{mf} CV_W^f)] = -4.1 \times 10^{-5}$ standard deviations per year.

According to our results, male collared flycatchers do, at least partly, keep their part of the bargain, whereas female collared flycatchers do not. The high heritability of the male ornament is in accordance with other estimates from wild populations²², and means that females can generally use ornaments to predict the sexual attractiveness of their offspring. The expression of the ornament was only weakly positively related to fitness, which is in-line with the theoretical expectation of low heritability of fitness in wild populations²³. In collared flycatchers, the high non-genetic component of variation in mate choice hampers the build-up of a covariance between mate choice and ornament, which in turn rules out the potential for indirect sexual selection. Instead, mate choice probably evolves in response to direct forces of selection because there are direct fecundity consequences for a female flycatcher's choice of mate with respect to the size of male ornament¹⁹. These findings illustrate the importance of testing the whole quantitative model underlying indirect sexual selection, and are consistent with the theoretical suggestion that indirect sexual selection on mate choice should be a weak force in absolute terms¹⁴. In species with other social settings (for example, lek-breeding species²⁴), comparing male phenotypes might be easier for females, which may result in heritabilities of mate choice that are closer to those estimates obtained in laboratory studies in which possible constraints of choice are controlled for²⁵. The high mating skew among males in lekking species may also result in higher variation in male fitness and a stronger correlation between male ornamentation and fitness. However, other quantities may be lower in such mating systems. For example, the pronounced sexual differences in life history in such species may reduce the overall coefficient of genetic variation across sexes, as illustrated by the lekking red deer, for which there is no additive genetic variance in female fitness²⁶. Thus, more empirical studies of all genetic parameters are needed before we can draw any clear conclusions about the strength of indirect sexual selection across different types of mating systems.

The low heritability of female mate choice in flycatchers is consistent with previous low estimates of repeatability (that is, the upper limit to heritability) of choice in the wild²⁵. This large component of non-genetic variation of mate choice indicates that in general, there are strong effects of a plethora of ecological interactions (for example, small-scale differences in mate availability, changes in environmental conditions and female–female competition²⁷) on the phenotypic expression of this trait. Moreover, the flexibility of a behavioural trait enables mate choice to change adaptively in accordance with such interactions^{19,28,29}. Therefore, environmental influences of mate choice are likely to be substantial in most natural systems, which markedly reduces the potential for indirect selection of mate choice. Future challenges for studies on the evolution of mate choice are to disentangle the relative role of ecological constraints versus adaptive changes in mate choice, and to investigate whether there is heritable variation in how females adjust their sampling and choice criteria in relation to changes in external and internal conditions.

Table 1 | Parameterization of components underlying models of indirect sexual selection in a natural population

Trait	Mean $\pm$ s.d.	V_R (s.e.m.)	V_A (s.e.m.)	V_{perm} (s.e.m.)	V_{plot} (s.e.m.)	V_{year} (s.e.m.)	CV_A
Ornament	76.3 $\pm$ 15.4	81.2 (0.10)	88.7 (0.43)	36.7 (0.39)	3.57 (0.07)	23.0 (0.32)	12.4
Choice	76.4 $\pm$ 15.3	202.4 (0.22)	6.24 (0.15)	0.000 (0.000)	3.55 (0.067)	26.0 (0.35)	3.3
Male fitness	0.65 $\pm$ 0.64	0.383 (0.013)	0.013 (7.9×10^{-3})	0.000 (2.3×10^{-3})	0.004 (2.6×10^{-3})	0.017 (6.3×10^{-3})	17.5
Female fitness	0.61 $\pm$ 0.64	0.367 (0.010)	0.018 (7.9×10^{-3})	0.008 (0.011)	0.002 (1.5×10^{-3})	0.012 (4.5×10^{-3})	22.0

Table shows the mean ($\pm$ s.d.) and the components of variance (s.e.m. in parentheses) in the male's ornament (forehead patch size), the female's observed mate choice for this ornament and annual fitness for males and females. Variance was partitioned using an animal model into residual effects (V_R), additive genetic effects (V_A), permanent environment (V_{perm}) and year effects (V_{year}). The coefficient of variation (CV, expressed as a percentage of the mean, $100 \times \sqrt{\text{variance}/\text{mean}}$) for the additive genetic component (CV_A) is a standardized measure of variance. Statistically significant components ($P < 0.05$) are indicated in bold. The genetic correlation between male and female annual fitness was positive but not significant ($r_W^{mf} = 0.234 \pm 0.354$, $P = 0.51$). Genetic correlations between ornament and choice (r_{CO}) and ornament and annual fitness (r_{OW}) are given in the main text.

METHODS

Data set. We used data obtained between 1980 and 2003 on ringed collared flycatchers (*Ficedula albicollis*) that bred in boxes in a study population consisting of a number of forest patches (plots) on the Swedish island of Gotland (57° 10' N, 18° 20' E). In total, the data set on forehead patch size and mate choice consisted of 4,220 males (6,803 measurements of forehead patch) and 4,345 females (6,338 measurements of mate choice). An individual's annual fitness is an estimate of its genetic contribution to next year's breeding population, and is calculated as the sum of its own apparent survival plus half the number of offspring (both sexes, including polygynous matings) that are recruited back into the breeding population in the following year(s). The factor 0.50 is incorporated to account for parent-offspring relatedness. Data on annual fitness (4,459 observations on 3,019 males and 5,731 observations on 2,934 females) was calculated from 1980–2002. Annual fitness allows us to omit all observations on broods for which a life-history experiment was conducted (for example, manipulation of brood size, cross-fostering; see Supplementary Information for details on data restriction). In this respect, we view annual fitness as a less biased estimate of fitness than lifetime reproductive success, for which either exclusion or (partial) inclusion of experimental broods will cause bias. This estimate of fitness is conservative with respect to the correlation between fitness and forehead patch, because it ignores extra-pair paternity (about 15% of nestlings), which is known to covary positively with forehead patch size¹⁸.

Animal model analysis. We used an 'animal model' (that is, a restricted maximum likelihood (REML) mixed model^{2,3}) to partition the variance in male fitness, male forehead patch size and female mate choice. As random effects for all traits, we incorporated additive genetic, permanent environment (that is, individual-specific), study plot and year effects. See Supplementary Information for details regarding data restrictions and which fixed effects were incorporated. Pedigree information was based on 3,715 individuals that recruited back into the breeding population and for which at least one genetic parent was known. Estimates of variance components are based on single-trait analysis, except for when calculating the genetic correlation between traits. The effect of erroneous paternity links in the pedigree due to unidentified extra-pair paternities on variance component estimates is probably negligible³⁰.

Received 16 November 2005; accepted 3 January 2006.

- Kokko, H., Brooks, R., Jennions, M. D. & Morley, J. The evolution of mate choice and mating biases. *Proc. R. Soc. Lond. B* **270**, 653–664 (2003).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, Massachusetts, 1996).
- Kruuk, L. E. B. Estimating genetic parameters in natural populations using the 'animal model'. *Phil. Trans. R. Soc. Lond. B* **359**, 873–890 (2004).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon Press, London, 1930).
- Lande, R. Models of speciation by sexual selection on polygenic traits. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725 (1981).
- Zahavi, A. Mate selection: a selection for a handicap. *J. Theor. Biol.* **53**, 205–214 (1975).
- Iwasa, Y., Pomiankowski, A. & Nee, S. The evolution of costly mate preferences. II. The handicap principle. *Evolution* **45**, 1431–1442 (1991).
- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
- Bakker, T. C. M. Positive genetic correlation between female preference and preferred male ornament in sticklebacks. *Nature* **363**, 255–257 (1993).
- Iyengar, V. K., Reeve, H. K. & Eisner, T. Paternal inheritance of a female moth's mating preference: consequences for sexual selection. *Nature* **419**, 830–832 (2002).
- Houde, A. E. Effect on artificial selection on male colour patterns on mating preference of female guppies. *Proc. R. Soc. Lond. B* **256**, 125–130 (1994).
- Brooks, R. & Coulridge, V. Multiple sexual ornaments coevolve with multiple mating preferences. *Am. Nat.* **154**, 37–45 (1999).
- Hine, E., Lachish, S., Higgie, M. & Blows, M. W. Positive genetic correlation between female preference and offspring fitness. *Proc. R. Soc. Lond. B* **269**, 2215–2219 (2002).
- Kirkpatrick, M. & Barton, N. H. The strength of indirect selection on female mating preferences. *Proc. Natl Acad. Sci. USA* **94**, 1282–1286 (1997).
- Service, P. M. & Rose, M. R. Genetic covariation among life history components: the effect of novel environments. *Evolution* **39**, 943–945 (1985).
- Pärt, T. & Qvarnström, A. Badge size in collared flycatchers predicts outcome of male competition over territories. *Anim. Behav.* **54**, 893–899 (1997).
- Qvarnström, A. Experimentally increased badge size increases male competition and reduces male parental care in the collared flycatcher. *Proc. R. Soc. Lond. B* **264**, 1225–1231 (1997).
- Sheldon, B. C., Merilä, J., Qvarnström, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character. *Proc. R. Soc. Lond. B* **264**, 297–302 (1997).
- Qvarnström, A., Pärt, T. & Sheldon, B. C. Adaptive plasticity in mate preference linked to differences in reproductive effort. *Nature* **405**, 344–347 (2000).
- Qvarnström, A. Genotype-by-environment interactions in the determination of the size of a secondary sexual character in the collared flycatcher (*Ficedula albicollis*). *Evolution* **53**, 1564–1572 (1999).
- Garant, D., Sheldon, B. C. & Gustafsson, L. Climatic and temporal effects on the expression of secondary sexual characters: genetic and environmental components. *Evolution* **58**, 634–644 (2004).
- Pomiankowski, A. & Møller, A. P. A resolution of the lek paradox. *Proc. R. Soc. Lond. B* **260**, 21–29 (1995).
- Burt, A. Perspective: the evolution of fitness. *Evolution* **49**, 1–8 (1995).
- Höglund, J. & Alatalo, R. V. *Leks* (Princeton Univ. Press, Princeton, 1995).
- Bakker, T. C. M. The study of intersexual selection using quantitative genetics. *Behaviour* **136**, 1237–1266 (1999).
- Kruuk, L. E. B. et al. Heritability of fitness in a wild mammal population. *Proc. Natl Acad. Sci. USA* **97**, 698–703 (2000).
- Dale, S., Rinden, H. & Slagsvold, T. Competition for a mate restricts mate search of female flycatchers. *Behav. Ecol. Sociobiol.* **30**, 165–176 (1992).
- Lesna, I. & Sabelis, M. W. Diet-dependent female choice for males with 'good genes' in a soil predatory mite. *Nature* **401**, 581–584 (1999).
- Hunt, J., Brooks, R. & Jennions, M. D. Female mate choice as a condition-dependent life-history trait. *Am. Nat.* **166**, 79–92 (2005).
- Charmantier, A. & Reale, D. How do misassigned paternities affect the estimation of heritabilities in the wild? *Mol. Ecol.* **14**, 2839–2850 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Berglund, M. Björklund, M. Kirkpatrick and T. Pärt for comments on the manuscript and for technical advice, and all people involved in field work over the years. Financial support was provided by the Swedish Research Council (to A.Q. and L.G.) and the Academy of Finland (to J.E.B.).

Author Contributions A.Q. came up with the idea and wrote the manuscript together with J.E.B., who also did all the analyses. L.G. organized the long-term study, and all three authors took active part in discussion and commenting on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.Q. (anna.qvarnstrom@ebc.uu.se).

A distal enhancer and an ultraconserved exon are derived from a novel retroposon

Gill Bejerano¹, Craig B. Lowe¹, Nadav Ahituv^{2,3}, Bryan King^{1,4}, Adam Siepel^{1,†}, Sofie R. Salama^{1,4}, Edward M. Rubin^{2,3}, W. James Kent¹ & David Haussler^{1,4}

Hundreds of highly conserved distal *cis*-regulatory elements have been characterized so far in vertebrate genomes¹. Many thousands more are predicted on the basis of comparative genomics^{2,3}. However, in stark contrast to the genes that they regulate, in invertebrates virtually none of these regions can be traced by using sequence similarity, leaving their evolutionary origins obscure. Here we show that a class of conserved, primarily non-coding regions in tetrapods originated from a previously unknown short interspersed repetitive element (SINE) retroposon family that was active in the Sarcopterygii (lobe-finned fishes and terrestrial vertebrates) in the Silurian period at least 410 million years ago (ref. 4), and seems to be recently active in the 'living fossil' Indonesian coelacanth, *Latimeria menadoensis*. Using a mouse enhancer assay we show that one copy, 0.5 million bases from the neuro-developmental gene *ISL1*, is an enhancer that recapitulates multiple aspects of *Isl1* expression patterns. Several other copies represent new, possibly regulatory, alternatively spliced exons in the middle of pre-existing Sarcopterygian genes. One of these, a more than 200-base-pair ultraconserved region⁵, 100% identical in mammals, and 80% identical to the coelacanth SINE, contains a 31-amino-acid-residue alternatively spliced exon of the messenger RNA processing gene *PCBP2* (ref. 6). These add to a growing list of examples⁷ in which relics of transposable elements have acquired a function that serves their host, a process termed 'exaptation'⁸, and provide an origin for at least some of the many highly conserved vertebrate-specific genomic sequences.

One of the most evolutionarily constrained regions in mammalian genomes is the ultraconserved element uc.338 (ref. 5), a mammal-specific 223-base-pair (bp) region perfectly conserved between human, mouse and rat, overlapping a short protein-coding exon of *PCBP2* (ref. 6). This small region was observed to have multiple paralogues within the human genome, overlapping protein-coding exons of otherwise unrelated genes, as well as conserved intronic and intergenic regions⁹ (Supplementary Fig. S1). This region also has multiple homologues in coelacanth that are closer in sequence to the human ultraconserved element than many of its human paralogues (Fig. 1c).

Further scrutiny of the 1 million bases (Mb) of sequence available from the Indonesian coelacanth reveals that the match is contained in a 481-bp genomic repeat. A total of 59 closely related copies are found in all four different coelacanth genomic regions sequenced so far (Supplementary Table S1). Using these copies we reconstructed a consensus coelacanth sequence of this repeat (Supplementary Fig. S2). Despite the huge evolutionary separation between the two species, the coelacanth repeat consensus is 80% identical over 360 bp to the human region containing the ultraconserved element. We could find no significant similarity between the coelacanth sequence

and any known repeat. However, SINE families are often generated by the fortuitous retroposition of a transfer RNA sequence into a location where it can provide a DNA polymerase III (Pol III) promoter for a retrotranscriptionally capable transcript^{10,11}. Indeed, the coelacanth 5' end is similar to the vertebrate serine tRNA, conserving the A and B boxes of an internal Pol III promoter, the 3' end has a clear poly(A) region, and the sequence is free of internal oligothymidylate tracts (Fig. 1a). This, combined with the high copy number (extrapolated to about 10⁵ copies genome-wide), low divergence between copies and evidence of target site duplications, indicates that the *L. menadoensis* sequences define a recently active SINE family, which we term the LF-SINE, for lobe-finned fishes (or 'living fossil') SINE. This family shares a weak 65-bp signature with two superfamilies of known SINEs (Supplementary Information S1).

Diverged LF-SINE copies are found in all available tetrapod genome drafts (Supplementary Table S3), as well as among the partial genomic data from a related coelacanth species and from multiple amniotes (Supplementary Table S4). We cannot detect significant LF-SINE matches in lungfish DNA sequence currently available (under 300 kilobases (kb)) or in genome drafts of available ray-finned fish or invertebrates (Supplementary Table S2). Nor is there any sequence similarity evidence in the public repositories indicating possible non-hereditary (horizontal) transfer from any other DNA source (Supplementary Information S2). It therefore seems that this SINE family was generated by a tRNA retroposition in a species of the ancestral Sarcopterygii and is specific to this clade (Fig. 2).

There are only several hundred recognizable copies of the LF-SINE in tetrapods for which we have genome drafts: 245 in human, 235 in dog, 394 in opossum, 699 in chicken and 26 in frog (Supplementary Table S3). These copies form orthology groups, in which each orthologue is in the same relative location with respect to the surrounding genes in all tetrapods where it is present (Fig. 1c). Each such group represents a single LF-SINE retroposition that occurred before the common ancestor of the species in the group. As expected from a recently active SINE, none of the 59 coelacanth instances have a human orthologue. However, multiple instances in tetrapods, including 29 in human (Supplementary Fig. S15), match the entire span of the reconstructed coelacanth SINE, including portions of its poly(A) tail, providing direct evidence for retroposition activity in the tetrapod lineage. This analysis establishes that virtually all retroposition events that generated mammalian LF-SINE instances predate the divergence of placental mammals and marsupials and that at least several of them, possibly all, predate the divergence of amniotes and amphibians. Examination of orthology groups using a very conservative test indicates that most human instances and their orthologues have evolved significantly more

¹Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. ²DOE Joint Genome Institute, Walnut Creek, California 94598, USA. ³Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁴Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. [†]Present address: Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA.

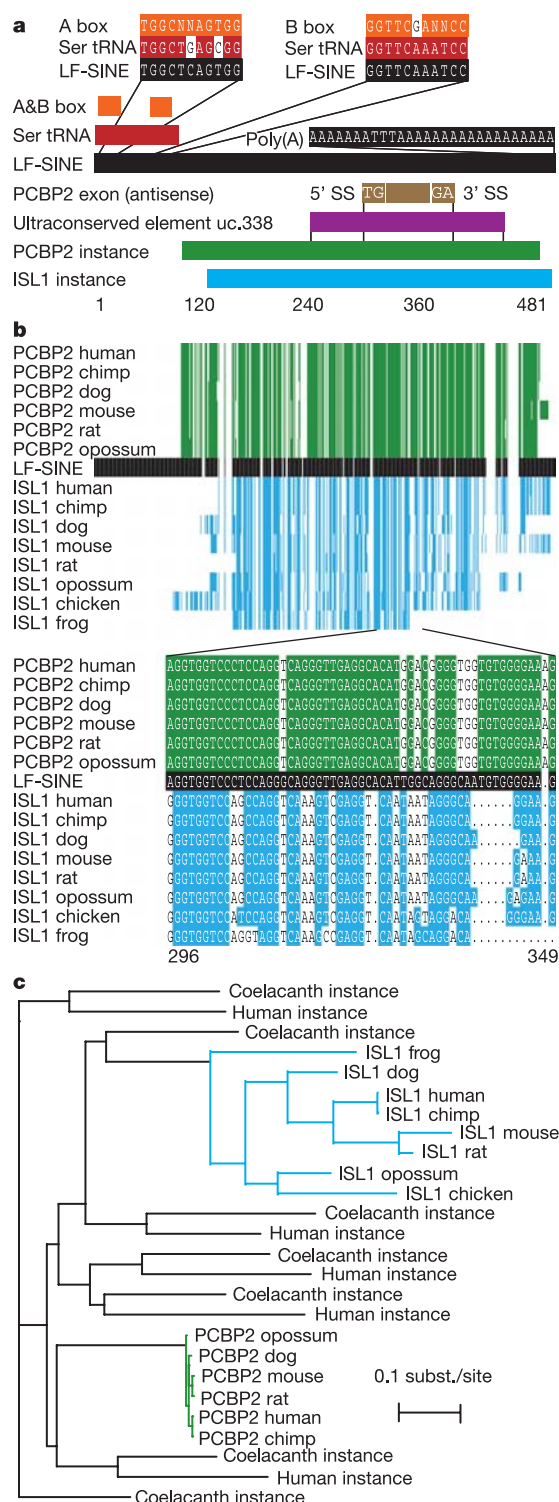


Figure 1 | Coelacanth SINE, human ultraconserved *PCBP2* exon and *ISL1* proximal enhancer share a common origin. **a**, Anatomy of the LF-SINE and its relation to an exapted tetrapodal distal enhancer near *ISL1*, and the ultraconserved exon of *PCBP2*, exonized from the reverse strand. SS, splice site. **b**, Alignment of multiple species instances of the *PCBP2* exonized element, and *ISL1* proximal LF-SINE enhancer, with the reconstructed coelacanth SINE. Filled squares (matches) and white spaces (tetrapodal inserts) are with respect to the coelacanth sequence. **c**, A maximum-likelihood joint phylogeny of selected LF-SINE instances from multiple species. The orthologous copies are shown to form monophyletic subtrees, whereas the additional instances serve to demonstrate the remarkable overall similarity between human and coelacanth instances.

slowly than would be expected assuming neutrality (Supplementary Information S3). This indicates that most detectable instances of the LF-SINE in tetrapods might have been exapted into cellular roles benefiting the host, subjecting them to purifying selection. In some cases the exapted tetrapod instance is remarkably close to the coelacanth SINE, indicating that the active LF-SINE in coelacanth might have changed very little over more than 410 Myr of independent evolution. The dispersion of coelacanth instances over many subclades in the evolutionary tree for these elements (Fig. 1c) precludes the possibility of recent horizontal transfer from tetrapods to coelacanth.

Most human instances of LF-SINEs are either intergenic (163 of 245; 66%; 107 more than 100 kb from a known gene) or intronic (68; 28%), and a smaller subset (14; 6%) overlap documented exons. We cannot find transcriptional evidence or predictions indicating that the human LF-SINEs are active as small RNAs or are involved in antisense regulatory transcripts. However, LF-SINE instances are found preferentially near genes involved in transcriptional regulation and neuronal development, indicating possible exaptation to form distal *cis*-regulatory regions (Supplementary Information S4).

To test this hypothesis, we picked a likely enhancer candidate and tested it *in vivo* using mouse transient transgenics. The *ISL1* gene encodes a LIM homeobox transcription factor that is required for motor neuron differentiation¹² and is expressed in motor and sensory neurons during vertebrate embryogenesis¹³. An *ISL1* proximal LF-SINE instance, significantly conserved between mammals, chicken and frog, lies 488 kb downstream of *ISL1*, in a 1.4-Mb gene desert that is home to two confirmed distal enhancers¹³ (Fig. 3a). The relative ordering and proximity to *ISL1* of the previously characterized enhancers and the LF-SINE instance represent an ancient organization that is invariant in frog, chicken, opossum, mouse and human (Supplementary Fig. S8).

The human *ISL1* proximal LF-SINE instance was cloned upstream of a mouse minimal heat shock 68 (*Hsp68*) promoter coupled to the β -galactosidase (*lacZ*) reporter gene and injected into the pronuclei

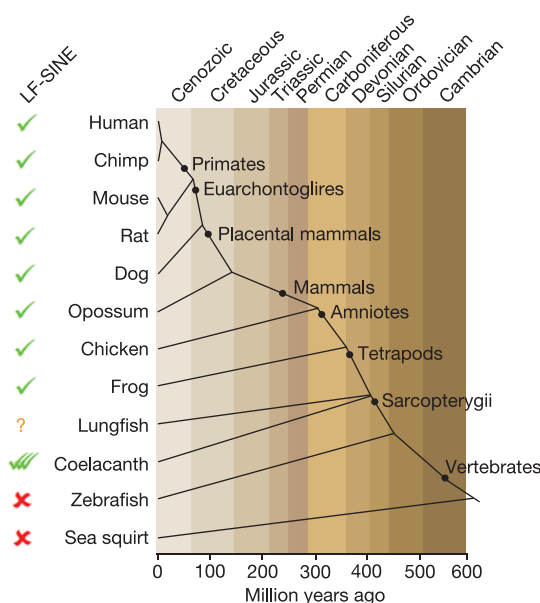


Figure 2 | Phylogeny of chordate genomes searched for instances of the LF-SINE. LF-SINE copies were found in the draft genomes of all terrestrial vertebrates shown and in genomic regions available from two coelacanth species. The LF-SINE was not found in very partial genomic data from lungfish, nor in any available draft genome of non-sarcopterygian vertebrates and invertebrates, including the two shown here. Temporal estimates are taken from ref. 4 and later sources. One tick, 25–700 copies in genome draft; three ticks, 59 copies in 1 Mb of DNA; question mark, no copies in less than 300 kb of DNA; cross, no copies in genome draft.

of fertilized mouse oocytes. The resulting embryos were analysed at embryonic day 11.5 (E11.5) by whole-embryo staining for *lacZ* activity (see Methods). Eight of nine independent *ISL1* proximal LF-SINE transgenic embryos showed consistent expression in the head and spinal cord region, the dorsal apical ectodermal ridge and genital eminence; in addition, four of nine embryos showed staining in the trigeminal ganglion (Fig. 3). Horizontal sections demonstrate specific colocalization of the *ISL1* proximal LF-SINE-driven *lacZ* reporter and murine *Isl1* RNA in neural tissues (Fig. 4). These expression patterns clearly recapitulate aspects of *Isl1* expression in developing motor neurons at this developmental stage^{13,14}. The novel and the two previously described enhancers in this region drive a very similar pattern of reporter gene expression at E11.5. They may drive expression distinctively at a different time point, perhaps later in development, as data for the two known enhancers seem to indicate¹³. Our combined functional and evolutionary analysis indicates that this LF-SINE instance might have been exapted as an *ISL1* enhancer before the divergence of the tetrapods and still functions in this capacity today. This constitutes a proof that mobile elements give birth to distal enhancers.

The *ISL1* proximal LF-SINE instance and the instance overlapping ultraconserved region uc.338 have conserved a very similar portion of the ancient LF-SINE (Fig. 1). However, one serves as a distal enhancer, and the other as an alternatively spliced exon. To gain a better understanding of exonization, we examined all 19 LF-SINE

instances that were exapted into protein-coding mRNAs (Supplementary Table S6). The affected proteins, encoded by *PCBP2*, *SMARCA4*, *EEF1B2*, *TCERG1*, *PTDSR*, *RORA*, *GRID1*, *ATF2*, *FLJ22833*, *ARHGAP6*, *KIAA1409*, *NT5C2*, *LRP1B*, *DHX30*, *gg-DMTF1*, *gg-PPP2R2C*, *gg-SHFM1*, *xt-MBNL1* and *JGI-49280*, are unrelated. Only a single pair of them shares a structural domain (helicase). All 19 derived exons are antisense to the original LF-SINE transcript. In 17 of 19 cases a new exon is formed in the middle of the coding region. Only canonical splice sites are used, similarly yet distinct from primate specific Alu-SINE exonization¹⁵ (Supplementary Information S5). Exapted exons start in all three possible reading frames. Sixteen of 17 are alternatively spliced, potentially leaving the original functional isoforms intact while evolution optimized the function of the novel isoform¹⁶. Eleven of 17 introduce an early stop codon, predicted to trigger nonsense-mediated decay¹⁷. Often the most evolutionarily conserved regions are the LF-SINE-derived intronic regions immediately flanking the exons, indicating the possible presence of exapted regulatory elements. Taken together, these observations do not indicate a common protein structural modification induced by exonization of the LF-SINE. Rather, LF-SINE exaptation might be used to regulate the protein levels, including in *PCBP2*, in which the ultraconserved exon might be involved in cellular localization¹⁸, dimerization¹⁹ and post-transcriptional auto-regulation²⁰, as well as in *SMARCA4* (*BRG1*; ref. 21) and *LRP1B* (ref. 22; Supplementary Information S5).

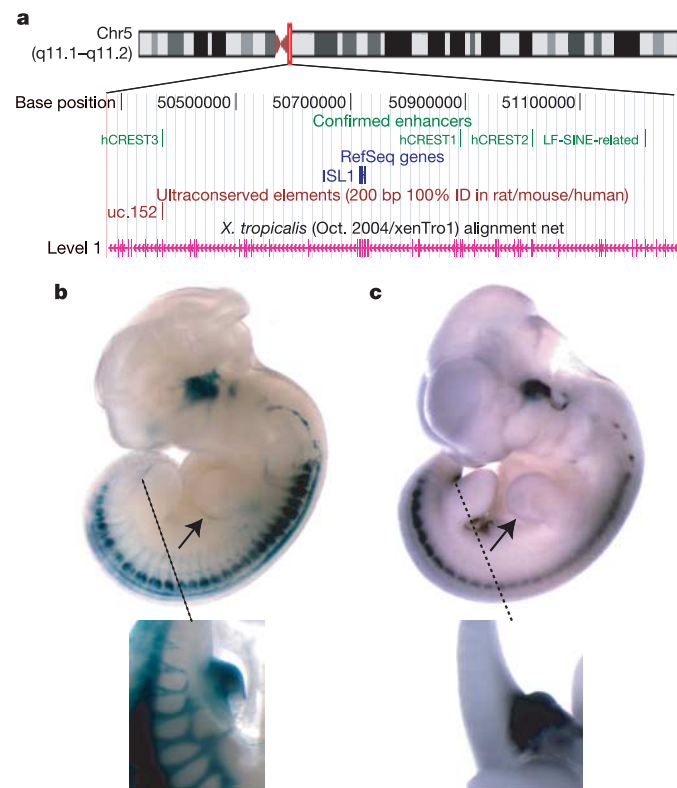


Figure 3 | A SINE-derived distal enhancer near *ISL1*. **a**, A 1-Mb pericentromeric neighbourhood of *ISL1* holds three previously confirmed enhancers¹³ (hCREST1, hCREST2 and hCREST3 = uc.152), and the novel LF-SINE-derived enhancer, 488 kb downstream of *ISL1*. The genomic organization of *ISL1* and the four enhancers is conserved between human and frog (*Xenopus tropicalis*). **b**, Expression pattern of a representative reporter gene construct driven by the human *ISL1* proximal LF-SINE in a transient transgenic mouse at E11.5. **c**, This pattern recapitulates major aspects of the expression pattern of the mouse *Isl1* gene at E11.5, assayed with whole-mount *in situ* hybridization. Enlargements show the genital eminence and arrows indicate the staining of the dorsal apical ectodermal ridge.

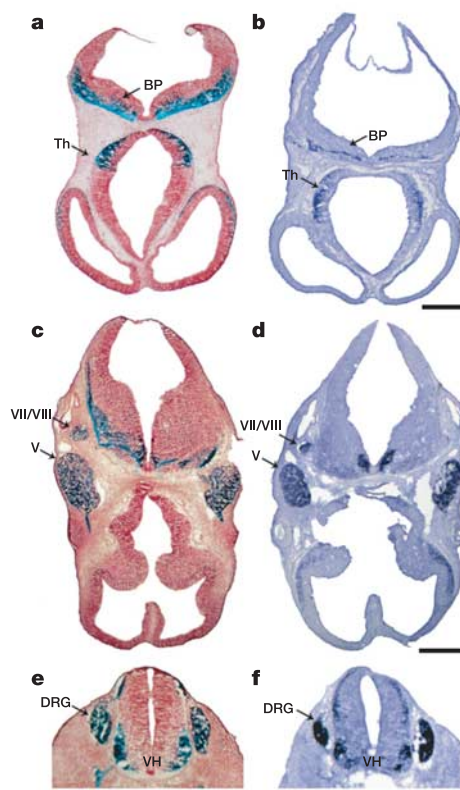


Figure 4 | Neural-specific expression driven by *ISL1*-proximal-LF-SINE recapitulates *Isl1* expression. Horizontal sections through E11.5 mice. **a, c, e**, *LacZ* staining in blue from the *ISL1*-LF-SINE-*LacZ* transient transgenic embryos with a neutral red counterstain. **b, d, f**, *In situ* RNA hybridization of *Isl1* in wild-type embryos. Matched level sections show corresponding expression patterns in the developing thalamus (Th) and basal plate (BP) in the brain (**a, b**), the trigeminal (V) and facio-acoustic (VII/VIII) ganglia in the head region (**c, d**), and the dorsal root ganglion (DRG) and the lateral region of the ventral horn (VH) of the spinal cord (**e, f**; thoracic sections). In **a-d** posterior is up; in **e** and **f** dorsal is up. Scale bars, 0.5 mm.

After discovering mobile DNA elements, Barbara McClintock suggested that they were fundamentally involved in gene regulation²³, an idea further developed by Britten and Davidson, who speculated on the benefit of obtaining similar control regions for a 'battery' of co-regulated genes through exaptation²⁴. At least 50% of our genome originates from characterized transposon-derived DNA²⁵. Although the early systematic theories of their role in gene regulation were not confirmed, it seems possible that, because these elements optimize their interaction with the host machinery under strong, virus-like evolutionary pressures, they are a particularly fecund source of evolutionary innovations, including new gene regulatory elements, and these are at times exapted by the host to improve its own fitness⁷. If so, it is possible that many more of the one million conserved vertebrate genomic elements originated from ancient retroposon families. In support of this hypothesis we find thousands of paralogue families of highly conserved non-coding sequences in the human genome^{9,26}, as well as individual exons with multiple non-coding paralogues (for example, Supplementary Fig. S12), of unknown origins (Supplementary Information S6).

The SINE families that are active in the eight tetrapods for which we have so far obtained draft genomes have all been restricted to specific clades, indicating rather recent origins and thus a rapid turnover of active SINE families on an evolutionary timescale. The Indonesian coelacanth may be different in that its LF-SINEs, and independently discovered Deu-SINEs (H. Nishihara, A. Smit and N. Okada, personal communication), have apparently remained active for more than 400 Myr with very little change. By preserving what in other species would be transient transposon families, the coelacanth acts, in a sense, as a living molecular fossil. The remaining 99.9% of its genome, as yet unsequenced²⁷, may very well hold precious traces of additional events that helped shape our own evolution.

METHODS

Enhancer analysis. The *ISL1* proximal LF-SINE instance was amplified from human DNA (BD Biosciences Clontech) by polymerase chain reaction (PCR) with the following primers: forward, 5'-AACATCTTGAAAAGAAGATCTAAGC-3'; reverse, 5'-AAGCTGCTTTTAAACTGTATCTTC-3'. The amplified DNA was cloned into the *Hsp68-lacZ* vector²⁸. The construct was then purified and injected into pronuclei as described previously²⁹. Embryos were harvested at E11.5, and transgenic embryos were identified by PCR of *lacZ*, using DNA from yolk sac and the following *lacZ* primers: forward, 5'-TTTCCATGTTGCCACTCGC-3'; reverse, 5'-AAGCGCTTGCCGTCAGCA-3'. Expression of *lacZ* was assayed in all embryos, with 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-Gal; Sigma), as described previously²⁹. See also (<http://enhancer.lbl.gov/aboutproject.html>). The *ISL1* proximal LF-SINE transient transgenics were further analysed by examining 20- μ m horizontal sections made from cryo-preserved embryos. Sections were counterstained with neutral red (Sigma; 0.3% w/v in PBS) to reveal the tissues. For *in situ* RNA hybridizations a murine *ISL1*-containing plasmid (IMAGE no. 3419119) was used to make sense and antisense digoxigenin-labelled RNA probes. *In situ* hybridizations on wild-type E11.5 embryos were performed as described previously³⁰. Stained sections were photographed with a PowerShot G6 digital camera (Canon) mounted on a dissecting microscope.

Computational analysis. The computational methods are described in Supplementary Information.

Received 27 November 2005; accepted 2 March 2006.

Published online 16 April 2006.

1. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
2. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
3. International Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
4. Graur, D. & Li, W.-H. *Fundamentals of Molecular Evolution* 2nd edn, Appendix I (Sinauer, Sunderland, Massachusetts, 2000).
5. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).

6. Makeyev, A. V. & Liebhaber, S. A. The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *RNA* **8**, 265–278 (2002).
7. Brosius, J. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* **118**, 99–116 (2003).
8. Gould, S. & Vrba, E. Exaptation—a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
9. Bejerano, G., Haussler, D. & Blanchette, M. Into the heart of darkness: Large-scale clustering of human non-coding DNA. *Bioinformatics* **20**, i40–i48 (2004).
10. Weiner, A. M. SINEs and LINEs: The art of biting the hand that feeds you. *Curr. Opin. Cell Biol.* **14**, 343–350 (2002).
11. Deininger, P. L. & Batzer, M. A. Mammalian retroelements. *Genome Res.* **12**, 1455–1465 (2002).
12. Pfaff, S. L., Mendelsohn, M., Stewart, C. L., Edlund, T. & Jessell, T. M. Requirement for LIM homeobox gene *Isl1* in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309–320 (1996).
13. Uemura, O. *et al.* Comparative functional genomics revealed conservation and diversification of three enhancers of the *Isl1* gene for motor and sensory neuron-specific expression. *Dev. Biol.* **278**, 587–606 (2005).
14. Caton, A. *et al.* The branchial arches and HGF are growth-promoting and chemoattractant for cranial motor axons. *Development* **127**, 1751–1766 (2000).
15. Lev-Maor, G., Sorek, R., Shomron, N. & Ast, G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* **300**, 1288–1291 (2003).
16. Makalowski, W. Genomics. Not junk after all. *Science* **300**, 1246–1247 (2003).
17. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
18. Chkheidze, A. N. & Liebhaber, S. A. A novel set of nuclear localization signals determine distributions of the alphaCP RNA-binding proteins. *Mol. Cell. Biol.* **23**, 8405–8415 (2003).
19. Kim, J. H., Hahn, B., Kim, Y. K., Choi, M. & Jang, S. K. Protein–protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **298**, 395–405 (2000).
20. Waggoner, S. A. & Liebhaber, S. A. Identification of mRNAs associated with α CP2-containing RNP complexes. *Mol. Cell. Biol.* **23**, 7055–7067 (2003).
21. Gunduz, E. *et al.* Genetic and epigenetic alterations of BRG1 promote oral cancer development. *Int. J. Oncol.* **26**, 201–210 (2005).
22. Li, Y., Lu, W. & Bu, G. Striking differences of LDL receptor-related protein 1B expression in mouse and human. *Biochem. Biophys. Res. Commun.* **333**, 868–873 (2005).
23. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA* **36**, 344–355 (1950).
24. Britten, R. J. & Davidson, E. H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46**, 111–138 (1971).
25. Kazazian, H. H. Jr. Mobile elements: Drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
26. McEwen, G. K. *et al.* Ancient duplicated conserved noncoding elements in vertebrates: A genomic and functional analysis. *Genome Res.* doi:10.1101/gr.4143406 (2006).
27. Danke, J. *et al.* Genome resource for the Indonesian coelacanth, *Latimeria menadoensis*. *J. Exp. Zool.* **301**, 228–234 (2004).
28. Kothary, R. *et al.* A transgene containing *lacZ* inserted into the dystonia locus is expressed in neural tube. *Nature* **335**, 435–437 (1988).
29. Poulin, F. *et al.* *In vivo* characterization of a vertebrate ultraconserved enhancer. *Genomics* **85**, 774–781 (2005).
30. Feldheim, D. A. *et al.* Topographic guidance labels in a sensory projection to the forebrain. *Neuron* **21**, 1303–1313 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the various sequencing consortia and research groups for the numerous genomic regions used in this study; A. Hinrichs, M. Diekhans, R. Harte, G. Barber and the UCSC browser team, M. Shoukry, I. Plajzer-Frick, S. Chanan and V. Afzal for technical help; R. Baertsch, T. Furey, E. Margulies and J. S. Pedersen for sharing unpublished data; and W. Miller, M. Blanchette, D. Feldheim, M. Nobrega, M. Ares, C. Sugnet, M. Dermizakis and J. Brosius for discussions. Research conducted at University of California Santa Cruz was supported by the National Human Genome Research Institute and the Howard Hughes Medical Institute; Research conducted at the E.O. Lawrence Berkeley National Laboratory was supported by grants from the Programs for Genomic Application, the NHLBI and performed under a Department of Energy Contract, University of California.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.B. (jill@soe.ucsc.edu).

A small-molecule screen in *C. elegans* yields a new calcium channel antagonist

Trevor C. Y. Kwok^{1*}, Nicole Ricker^{1*}, Regina Fraser², Allen W. Chan⁴, Andrew Burns¹, Elise F. Stanley⁴, Peter McCourt^{2,3}, Sean R. Cutler² & Peter J. Roy^{1,3}

Small-molecule inhibitors of protein function are powerful tools for biological analysis¹ and can lead to the development of new drugs. However, a major bottleneck in generating useful small-molecule tools is target identification. Here we show that *Caenorhabditis elegans* can provide a platform for both the discovery of new bioactive compounds and target identification. We screened 14,100 small molecules for bioactivity in wild-type worms and identified 308 compounds that induce a variety of phenotypes. One compound that we named nemadipine-A induces marked defects in morphology and egg-laying. Nemadipine-A resembles a class of widely prescribed anti-hypertension drugs called the 1,4-dihydropyridines (DHPs) that antagonize the α_1 -subunit of L-type calcium channels^{2,3}. Through a genetic suppressor screen, we identified *egl-19* as the sole candidate target of nemadipine-A, a conclusion that is supported by several additional lines of evidence. *egl-19* encodes the only L-type calcium channel α_1 -subunit in the *C. elegans* genome^{4,5}. We show that nemadipine-A can also antagonize vertebrate L-type calcium channels, demonstrating that worms and vertebrates share the orthologous protein target. Conversely, FDA-approved DHPs fail to elicit robust phenotypes, making nemadipine-A a unique tool to screen for genetic interactions with this important class of drugs. Finally, we demonstrate the utility of nemadipine-A by using it to reveal redundancy among three calcium channels in the egg-laying circuit. Our study demonstrates that *C. elegans* enables rapid identification of new small-molecule tools and their targets.

To identify new small-molecule tools for the biological analysis of *C. elegans*, we screened 14,100 small membrane-permeable compounds⁶ for the induction of defects in wild-type worms. We identified 308 compounds that induced a variety of phenotypes including slow growth, lethality, uncoordinated movement and morphological defects (Supplementary Tables 1 and 2). This high hit rate demonstrates that *C. elegans* is a suitable model for identification of new bioactive compounds, and is a similar hit rate to that described for other whole-organism screens⁷. To explore the utility of *C. elegans* for target identification, we focused on one compound that we named nemadipine-A, which induces slow population growth defects (Fig. 1, Table 1 and Supplementary Fig. 1), called a Gro phenotype, dramatic defects in morphology (Fig. 2), called a variable abnormal or Vab phenotype^{8,9}, and egg-laying defects (Fig. 3), called an Egl phenotype.

To identify candidate targets of nemadipine-A, we screened 180,000 randomly mutated wild-type genomes for dominant genetic suppressors of the nemadipine-A-induced Gro phenotype. We isolated 16 suppressors, five of which we made homozygous and

characterized further. Each of these five mutants suppress the Gro, Egl and Vab phenotypes induced by nemadipine-A and are tightly linked to *egl-19* (Table 2). The *egl-19* locus encodes the only L-type calcium channel α_1 -subunit in the *C. elegans* genome^{4,5} and is so named because reduction-of-function mutants (hypomorphs) are Egl (refs 4, 10). All five of the nemadipine-A suppressors harboured missense mutations in conserved regions of *egl-19* that form the pore of the channel (Table 2 and Supplementary Fig. 3). Furthermore, three mutated residues are in regions required for DHP interaction with mammalian L-type channels¹¹, and one allele (*tr69*) alters a residue required for DHP interaction in mammalian systems. Unlike the previously characterized *egl-19* hypermorphs^{4,5}, two of the *egl-19* suppressors (*tr69*, *tr71*) have no hypermorphic activity as they do not lay eggs constitutively and are not short (Table 2). This suggests that the *tr69* and *tr71* mutations disrupt nemadipine-A interaction with EGL-19 and do not simply upregulate channel activity. Together with the fact that all of the characterized nemadipine suppressors are mutations in *egl-19*, these results strongly suggest that EGL-19 is a target of nemadipine-A.

Four additional lines of genetic evidence suggest that EGL-19 is the primary target of nemadipine-A. First, wild-type (N2) worms raised on nemadipine-A phenocopy previously characterized *egl-19* loss-of-function mutants⁴. Like *egl-19* hypomorphs, nemadipine-A-treated worms are Egl (Fig. 3). In addition, N2 worms grown on nemadipine-A have a Vab phenotype similar to strong hypomorph alleles of *egl-19* (Fig. 2)⁴. Wild-type worms cannot survive on high concentrations of nemadipine-A, consistent with the essential role of EGL-19 (ref. 4) (Supplementary Fig. 1). There are only two other voltage-gated calcium channels in the worm, UNC-2 and CCA-1, and their null phenotypes^{12,13} are inconsistent with either being a target of nemadipine-A. Second, *egl-19* hypomorphs are hypersensitive to nemadipine-A and exhibit the Vab and Gro defects on 10- to 20-fold less nemadipine-A than that required to elicit a phenotype in

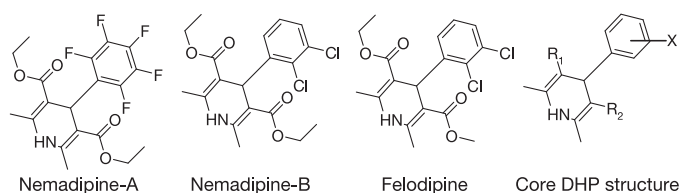


Figure 1 | The molecular structures of nemadipine-A, nemadipine-B, felodipine, and the core 1,4-dihydropyridine (DHP) structure. R₁, R₂ and X are defined in Table 1.

¹Department of Medical Genetics and Microbiology, and The Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. ²Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, Ontario M5S 3B2, Canada. ³Collaborative Program in Developmental Biology, University of Toronto, 25 Harbord Street, Toronto, Ontario M5S 3G5, Canada. ⁴Department of Physiology and Toronto Western Research Institute, University of Toronto, 399 Bathurst Street, Toronto, Ontario M5T 2S8, Canada.

*These authors contributed equally to this work.

N2 (Fig. 2 and Supplementary Fig. 1). Third, *egl-19* hypermorphs are resistant to nemadipine-A. The strong *egl-19* gain-of-function allele *n2368* (ref. 4) shows no Vab or Egl defects on all concentrations of nemadipine-A tested (Figs 2 and 3). The weaker *egl-19(ad695)* hypermorph⁴ exhibits Vab and Egl defects in nemadipine-A concentrations that are 10- to 50-fold greater than that which is required to elicit phenotype from N2 animals. Finally, nemadipine-A can rescue the defects resulting from EGL-19 hypermorphic activity (Table 2). The egg-laying constitutive phenotype, fecundity defects, and the short phenotype caused by prolonged muscle contraction in *egl-19(n2368)* and *egl-19(ad695)* hypermorphs^{4,5} are all rescued by nemadipine-A (Table 2, Fig. 3 and data not shown). Together with the facts that nemadipine-A is a DHP and that EGL-19 is an L-type calcium channel^{4,5}, these four additional lines of evidence strongly suggest that EGL-19 is the primary target of nemadipine-A.

None of the six FDA-approved DHPs from our libraries induced Egl or Vab phenotypes in our small-molecule screen, including amlodipine, felodipine, nifedipine, nimodipine and nitrendipine; nor did other classes of L-type calcium channel antagonists represented by verapamil, diltiazem and bepridil. The only phenotype observed with any of these was a mild Gro phenotype induced by felodipine (Table 1). These results were surprising because previous work demonstrated that both nifedipine and verapamil antagonize EGL-19 in dissected worms^{5,14,15}.

To investigate why nemadipine-A can antagonize EGL-19 in whole-worm assays better than other DHPs, we first explored the possibility that nemadipine-A is a relatively potent inhibitor of L-type calcium channels in general. We compared the ability of nemadipine-A and nifedipine to directly inhibit vertebrate L-type calcium channels in chick ciliary neurons¹⁶. Patch-clamp analysis showed that nemadipine-A and nifedipine inhibited L-type calcium channels at approximately the same potency (Supplementary Fig. 4). Previous work has demonstrated that the six FDA-approved DHPs have similar potencies¹⁷. This suggests that nemadipine-A activity in whole worms does not result from a more potent target-interaction relative to other DHPs.

We then tested the hypothesis that different DHP activity in whole worms results from differences in their accumulation in worms. High-performance liquid chromatography (HPLC) was used to determine the amount of compound in N2 worms after either a two-hour liquid incubation with the drugs, or after N2 worms were raised on solid substrate containing the antagonists for three days. Both approaches revealed that only nemadipine-A and felodipine accumulated in worms to a concentration similar to that present in the media (Table 1 and data not shown). These two molecules also demonstrated activity in the N2 population growth assays. However, only nemadipine-A exhibits robust EGL-19 antagonism in wild-type worms, demonstrating that the structure of nemadipine-A enables both accumulation in worms and robust EGL-19 antagonism.

We next investigated DHP structure-activity relationships to determine which substructures enable accumulation in worms and EGL-19 antagonism. Intriguingly, felodipine is most like nemadipine-A in structure and is the only therapeutic DHP that we have investigated that has activity in whole worms. This suggests that their shared poly-halogenated phenyl groups and/or ethyl ester substitutions may be responsible for efficacy in worms. We therefore examined analogues of nemadipine-A that either lacked fluorines or had a single fluorine on the phenyl group (analogues 1–4). None of these analogues accumulated extensively or induced phenotype in wild-type worms (Table 1). This demonstrates that multiple halogens on the DHP phenyl group are required for extensive accumulation in worms, which is a likely prerequisite for robust EGL-19 antagonism.

Nemadipine-A and felodipine also have ester groups in common, but differ in that nemadipine-A has 3,5-diethyl esters, while felodipine has 3-methyl 5-ethyl esters (Fig. 1). We therefore tested the hypothesis that the symmetrical ethyl esters of nemadipine-A confers robust EGL-19 antagonism by examining a 3,5-ethyl ester analogue

of felodipine. This analogue, which we renamed nemadipine-B, accumulated in worms to similar levels as nemadipine-A ($P > 0.05$) (Fig. 1 and Table 1). Nemadipine-B also induced extensive Egl, Vab and Gro phenotypes, consistent with EGL-19 antagonistic activity (Figs 2 and 3, and Table 1). By contrast, a dimethyl ester analogue of felodipine (analogue 6) neither accumulated extensively nor induced a phenotype. The position of the dichloro groups on nemadipine-B is probably required for EGL-19 antagonism because chloro groups in different positions (analogue 7) enabled accumulation, but induced no phenotype (Table 1). Together, our structure-activity analysis demonstrates that multiple halogens on the DHP phenyl ring and at least one ethyl ester are critical for extensive DHP accumulation in worms, but is insufficient for robust EGL-19 antagonism *in vivo*. Instead, DHP analogues with 3,5-ethyl esters and halogens in the *ortho* and *meta* positions on the phenyl ring both

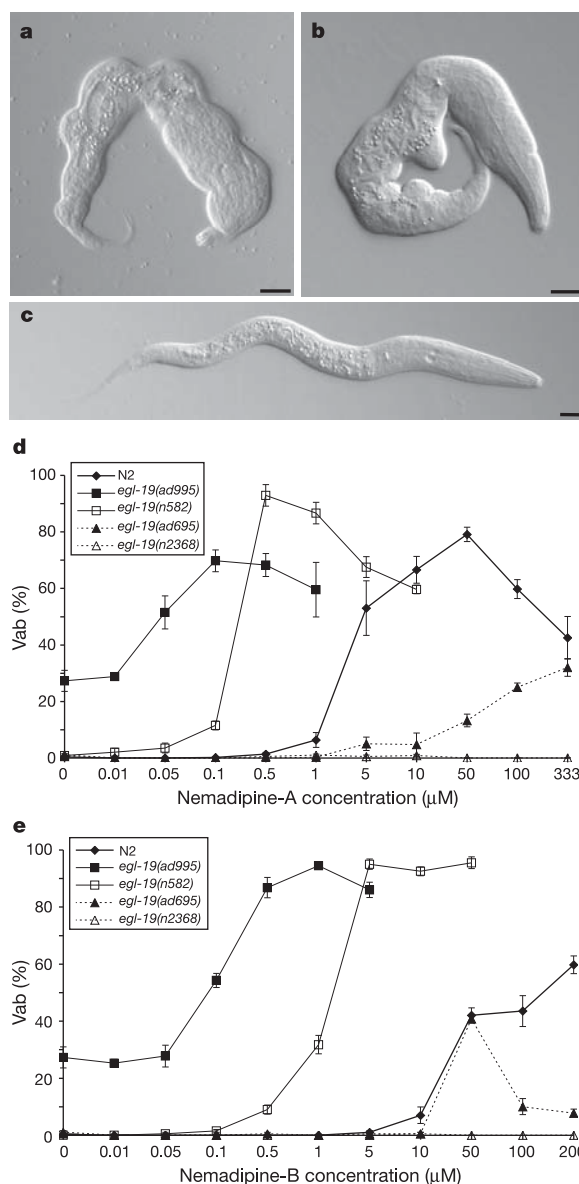


Figure 2 | Embryos raised on nemadipine hatch with Vab defects. **a**, An N2 hatchling raised on 10 μM nemadipine-A. **b**, An *egl-19(ad995)* hypomorph hatchling. **c**, An N2 hatchling raised on a DMSO control plate. Scale bars, 7.5 μm. **d**, **e**, Dose-response curves of the percentage of the indicated genotypes that hatch with Vab defects on nemadipine-A (**d**) and nemadipine-B (**e**). The graphs show the response to the limit of solubility for nemadipine-A (333 μM) and nemadipine-B (200 μM)¹. Error bars, $\pm$ s.e.m.

Table 1 | Biological properties of L-type calcium channel antagonists and DHP analogues

Compound*	R ₁ †	R ₂ †	X†	Egl (%)‡	Vab (%)§	Gro	Accumulation#
None	–	–	–	<2	<2	OG	–
Nemadipine-A	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	2,3,4,5,6-F ₅	93.0 ± 4.2	68.0 ± 6.4	1.5 ± 0.6	1.6 ± 0.5
Felodipine	CO ₂ CH ₃	CO ₂ CH ₂ CH ₃	2,3-Cl ₂	<2	<2	108.7 ± 31.6	1.1 ± 0.2
Nifedipine	CO ₂ CH ₃	CO ₂ CH ₃	2-NO ₂	<2	<2	OG	ND
Nimodipine	CO ₂ CH(CH ₃) ₂	CO ₂ CH ₂ CH ₂ OCH ₃	3-NO ₂	<2	<2	OG	0.2 ± 0.1
Nicardipine	CO ₂ CH ₃	CO ₂ CH ₂ CH ₂ NHCH ₂ C ₆ H ₅	3-NO ₂	<2	<2	OG	ND
Verapamil	–	–	–	<2	<2	OG	ND
Diltiazem	–	–	–	<2	<2	OG	ND
Bepridil	–	–	–	<2	<2	OG	0.1 ± 0.1
Analogue 1	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	–	<2	<2	OG	0.1 ± 0.0
Analogue 2	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	4-F	<2	<2	OG	0.4 ± 0.2
Analogue 3	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	3-F	<2	<2	OG	0.3 ± 0.2
Analogue 4	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	2-F	<2	<2	OG	0.2 ± 0.1
Nemadipine-B	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	2,3-Cl ₂	92.5 ± 4.4	40.3 ± 4.5	1.3 ± 0.5	1.2 ± 0.2
Analogue 6	CO ₂ CH ₃	CO ₂ CH ₃	2,3-Cl ₂	<2	<2	OG	0.2 ± 0.2
Analogue 7	CO ₂ CH ₂ CH ₃	CO ₂ CH ₂ CH ₃	3,4-Cl ₂	<2	<2	OG	1.4 ± 0.3

Errors shown are ±s.d.

* The structure, molecular weight, log*P* (where *P* is the partition coefficient), and commercial source of the compounds are shown in Supplementary Fig. 2.

† The chemical groups in the positions indicated in the core DHP structure in Fig. 1 are shown.

‡ The percentage of egg-laying defective N2 adults containing late-stage embryos.

§ The percentage of N2 hatchlings with morphological defects.

|| The number of L4s and adults produced by two N2 parents after 5 d. OG, wells are over-grown with progeny that have run out of food and burrowed into the agar substrate.

The amount of compound accumulated in worms during two hours of incubation in M9 solution and detected by HPLC analysis. The amount is normalized to the concentration of the compound used in the assay (25 μM). All assays were done with 25 μM of compound, except for the Egl and Vab assays that were also done to the limit of solubility for each of the FDA-approved drugs, but no phenotypes were observed. ND, not detectable.

accumulate and achieve robust L-type calcium channel antagonism in *C. elegans*.

Having established the nemadipines as unique reagents to antagonize EGL-19 in whole worms, we applied this tool to the analysis of egg-laying in worms. We asked if EGL-19 has redundant roles in egg-laying by compromising EGL-19 function with nemadipine in the background of other voltage-gated calcium channel mutants. Egg-laying by *C. elegans* occurs when the 16 vulval and uterine muscles contract, thereby opening the vulva and forcing embryos through it. Only two types of neurons directly synapse with the vulval muscles; the HSNs and the VCs, which also receive synaptic input from the HSNs^{18–20}. Serotonin release from the HSNs makes the vulval muscles competent to contract, while acetylcholine release from the HSNs and/or VCs probably triggers vulval muscle depolarization, contraction and egg-laying^{21,22}. EGL-19 is probably required in the vulval muscles for both competence and contraction²¹. Conversely, a second voltage-gated calcium channel α₁-subunit called UNC-2, which is of the N/P/Q-type²³, is required in the HSNs and/or VCs to negatively regulate egg-laying¹³. The only other voltage-gated calcium channel α₁-subunit encoded by the *C. elegans* genome is CCA-1, which is of the T-type and has a role in pharyngeal pumping^{12,23}.

To determine if EGL-19 has redundant roles in egg-laying, we tested if *unc-2* and *cca-1* null mutants retain eggs in low concentrations of nemadipine. Surprisingly, the nemadipines elicited

a synthetic Egl phenotype in both mutants (Fig. 4a). The synthetic Egl phenotypes were suppressed by the *egl-19(ad695)* hypermorph, further demonstrating that the interactions depend on EGL-19 hypomorphic activity. These results suggest that EGL-19 functions redundantly with both UNC-2 and CCA-1 to positively regulate egg-laying.

We next investigated the nature of the synthetic egg-laying phenotypes. Egg-laying defective mutants that cannot produce serotonin will lay their eggs on exposure to exogenous serotonin. Egl mutants that produce low levels of serotonin will release their eggs in the presence of exogenous imipramine, a serotonin uptake inhibitor, or exogenous serotonin. However, Egl mutants whose vulval muscles cannot respond to serotonin or cannot contract do not lay eggs in response to either serotonin or imipramine^{10,24,25}. *egl-19(ad995)* does not lay eggs in response to serotonin or imipramine (Fig. 4b, c), consistent with a role for EGL-19 in the vulval muscles²¹. Like *egl-19* hypomorphs, *unc-2* mutants grown on 0.1 μM nemadipine-A did not respond significantly to serotonin or imipramine (Fig. 4). In contrast, *cca-1* mutants grown on 0.1 μM nemadipine-A laid eggs in response to serotonin, but not imipramine. This suggests that CCA-1 and EGL-19 function redundantly in the release of serotonin. This redundant role for EGL-19 is consistent with previous reports of neuronal *egl-19* expression⁴. We speculate that the increase in egg-laying by *unc-2(e55)* mutants (grown without nemadipine) in imipramine but not serotonin is a result of increased excitability of

Table 2 | Genetic suppressors of nemadipine-induced defects

Allele	Mutation†	Reg. mut.	Egl (%) –nema	Egl (%) +nema	Vab (%) –nema	Vab (%) +nema	Length (mm)‡ –nema	Length (mm)‡ +nema	Embryos§ –nema	Embryos§ +nema
N2	Wild type	–	≤2	90.0 ± 6.7	≤2	53.0 ± 23.7	1.11 ± 0.04	1.12 ± 0.05	10.7 ± 2.3	31.3 ± 9.3
<i>ad695</i>	A906V	IIIS4	≤2	4.0 ± 1.4*	≤2	5.0 ± 4.8*	0.93 ± 0.07*	1.06 ± 0.05*	5.1 ± 2.2*	15.4 ± 5.7*
<i>n2368</i>	G365R	IS6	≤2	≤2*	≤2	≤2*	0.75 ± 0.05*	0.98 ± 0.06*	3.5 ± 1.8*	7.6 ± 2.5*
<i>tr69</i>	M1056I	IIIS6	≤2	≤2*	≤2	≤2*	1.10 ± 0.06	1.15 ± 0.04*	12.3 ± 3.0	17.6 ± 4.0*
<i>tr70</i>	S1010L	IIIS1-SS2	≤2	≤2*	≤2	≤2*	1.00 ± 0.04*	1.00 ± 0.04*	7.7 ± 2.0*	7.2 ± 2.3*
<i>tr71</i>	V1060L	IIIS6	≤2	36.5 ± 12.8*	≤2	≤2*	1.10 ± 0.06	1.15 ± 0.04*	13.9 ± 4.6	29.2 ± 6.1
<i>tr72</i>	I652F,V679I	IIIS1-SS2, IIS6	≤2	6.0 ± 4.0*	7.3 ± 7.6*	13.3 ± 5.3*	0.97 ± 0.05*	1.11 ± 0.05	16.5 ± 3.6	21.0 ± 5.9*
<i>tr73</i>	A615V	IIS5	≤2	≤2*	≤2	≤2*	1.04 ± 0.04*	1.16 ± 0.13*	12.4 ± 3.5	12.4 ± 5.0*

An asterisk indicates a significant difference from N2 (*P* < 0.001, Student's *t*-test). Reg. mut., region mutated; –nema, animals were raised on DMSO control plates; +nema, animals were raised on 5 μM nemadipine-A. Errors shown are ±s.d.

† The mutation notation is with respect to EGL-19 predicted peptide sequence (WormBase).

‡ The average length of 20 young adults.

§ The average number of embryos *in utero* of 20 young adults.

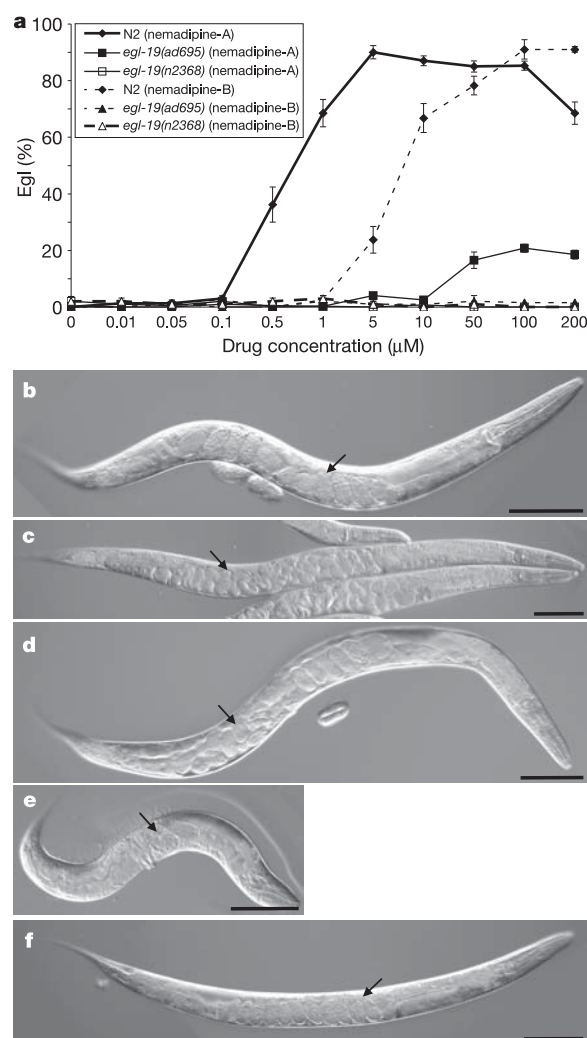


Figure 3 | Animals grown on the nemadipines are Egl. **a**, The percentage of animals that are Egl in response to either nemadipine-A or nemadipine-B. Error bars, $\pm$ s.e.m. **b**, An N2 adult raised on 0.33% DMSO with a normal number of embryos within the uterus (arrow). **c**, An N2 adult raised on 10 μM nemadipine-A containing many late-stage embryos (arrow). **d**, An *egl-19(n582)* hypomorphic adult that also contains several late-stage embryos. **e**, An *egl-19(n2368)* hypermorphic adult that is egg-laying constitutively and contains few embryos within the uterus (arrow) and is also myotonic, resulting in short body length. **f**, An *egl-19(n2368)* adult raised on 10 μM nemadipine-A. Scale bars, 100 μm.

the vulval muscles caused by imipramine antagonism of the EGL-2 potassium channel²⁶. Multiple calcium channel types may therefore be required in different cells of the egg-laying circuit for these cells to respond to voltage in a step-wise fashion. It is noteworthy that the redundancy of EGL-19, UNC-2 and CCA-1 would not be revealed without the ability to control EGL-19 activity with the nemadipines as the weakest *egl-19* hypomorph is already egg-laying defective⁴.

In summary, we have screened thousands of small molecules in a broad search for new bioactive compounds that might facilitate the study of *C. elegans* biology. The investigation of one of these bioactive compounds revealed a new L-type calcium channel antagonist that inhibits vertebrate channels and has unique properties that enable channel antagonism in whole worms. The nemadipines are reagents that now allow genetic investigation into DHP-target interactions *in vivo* and the identification of other loci that may modify the response to a class of molecules that are used by millions of people.

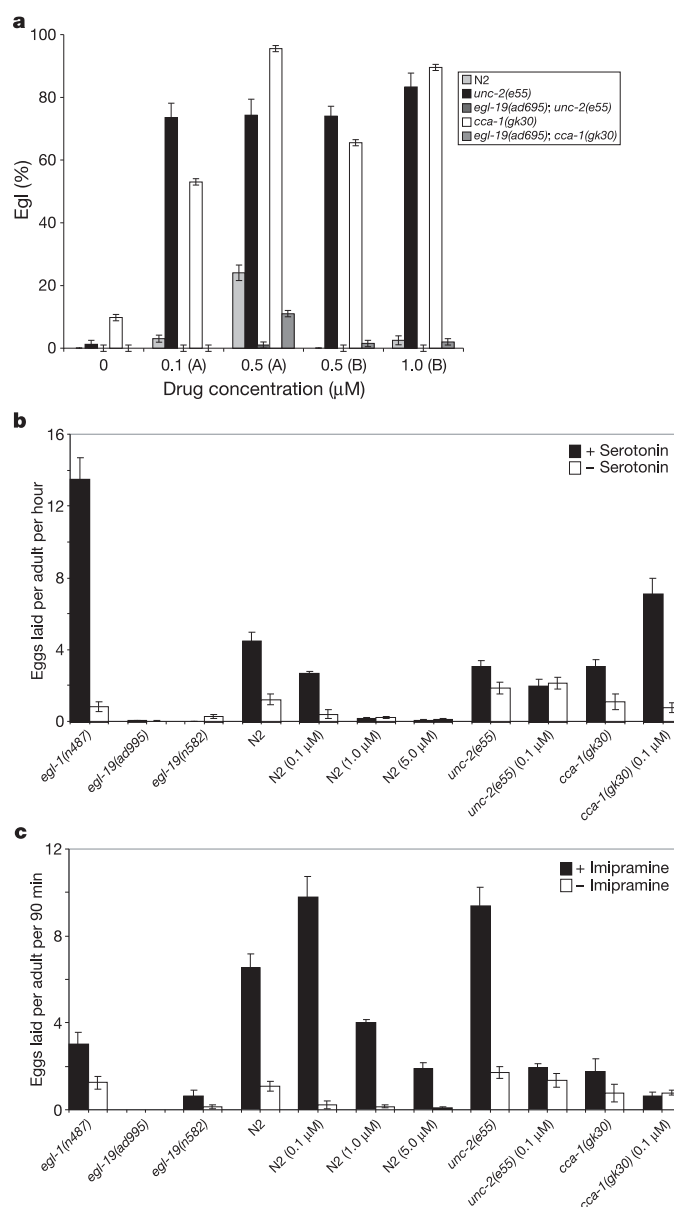


Figure 4 | EGL-19 functions redundantly with both UNC-2 and CCA-1. **a**, Nemadipine-treated *unc-2(e55)* and *cca-1(gk30)* null mutants retain late embryos. The *egl-19(ad695)* hypermorph suppresses these synthetic Egl phenotypes. Lower concentrations of the nemadipines produce little or no effect (data not shown) and higher concentrations produce a highly penetrant Egl phenotype in N2 worms (Fig. 2). The concentration is indicated (μM), followed by the drug used—A, nemadipine-A; B, nemadipine-B. **b**, **c**, The number of eggs laid in one hour in M9 liquid either with or without 5 mg ml⁻¹ serotonin (**b**), and the number of eggs laid in 90 min in M9 liquid either with or without 0.75 mg ml⁻¹ imipramine (**c**). The final concentration of nemadipine-A is indicated in brackets following the genotype. The assays were done as previously described^{10,24} (see Methods for more details). Error bars, $\pm$ s.e.m.

METHODS

Small-molecule screen. We identified 308 bioactive molecules by screening a total of four libraries of small molecules in duplicate from Chembridge Corp. (Diverset, 10,000 molecules), Microsource (Spectrum, 2,000 molecules), Sigma (LOPAC, 1,250 molecules) and Maybridge (Hits Kit, 1,000 molecules). This was done in 24-well plates with 1 ml of MYOB agar substrate²⁷ and 25 μl of stationary phase OP50 *Escherichia coli* on the surface of each well. The final concentration of each compound in each well was 25 μM assuming the molecules diffused throughout the media. A day after seeding the wells with *E. coli*, two wild-type third larval stage N2 worms were deposited into each well using a COPAS BIOSORT worm sorter (Union Biometrica). Three and four days later, each well

was scored independently with both a Leica MZFLIII microscope and from the micrographs generated by a HiDi 1.0 high-throughput digital imager (Elegenics Inc.).

Nemadipine suppressor screen, mapping and mutation identification. To screen for mutations that suppress nemadipine-A-induced defects, a mixed-stage population of wild-type N2 worms was mutagenized as previously described²⁸. Four days later, eight or twelve L3- or L4-staged F₁ larvae were deposited into each well of 24-well MYOB plates containing 10 or 12.5 μ M nemadipine-A using the COPAS sorter. Because N2 worms grow slowly on nemadipine-A, the wells were screened 6 to 7 days later for growth that obviously exceeded N2-containing wells. Twelve worms from these candidate wells were further tested for resistance on 10 or 12.5 μ M nemadipine-A. The progeny of faster-growing clones were tested for homozygosity. Of the 16 strains that showed consistent suppression, we were able to homozygote five quickly. See Supplementary Information for mapping details.

Phenotypic analysis. Vab and Egl assays were done by growing mixed-stage populations of worms on plates containing compounds at the desired concentrations or DMSO control. To ensure that Vab counts did not reflect an accumulation of arrested Vab hatchlings, adults were picked off each plate 24 h later and placed on a new plate containing the same compound or DMSO control and OP50 bacteria. Another 24 h later, the young L1 and L2 animals were scored for morphological defects. The Egl assay was done in the same manner as the Vab assay except that 24 h after transfer to a new plate with drug, the non-Egl and Egl adults were counted. Each experiment was done in quadruplicate.

For population growth assays, wells were seeded with two L3 worms. From 3–7 days after depositing the L3 parents, each well was photographed each day with either a Leica MZ7.5 and a PL-A661 PixeLINK camera or HiDi. The adults were then counted from the pictures. Each experiment was done in quadruplicate.

To measure the length of worms, ~100 L4s were first picked onto an MYOB plate with or without 5 μ M nemadipine-A and incubated overnight. Approximately 24 h later, pictures were taken of individual adults at 20 $\times$ magnification and the length of the animal was measured using Openlab software (Improvision, Inc.). The number of embryos *in utero* was counted in the same animals used for length measurements with a Leica MZFLIII microscope at ~80 $\times$ magnification.

Determination of drug concentration within worms. The accumulation of the small molecules within N2 wild-type worms was measured by growing synchronized hatchlings for two days at 20 $^{\circ}$ C (ref. 27). The fourth larval stage worms were harvested, rinsed at least twice, and then dispensed into aliquots of 50 μ l of packed worms (~7,700 worms). The worms were then incubated at room temperature for 2 h in 1 ml of M9 with 25 μ M of the compound, then rinsed 5 times on ice and the supernatant was removed. The samples were then stored at –20 $^{\circ}$ C and were later lysed by adding 50 μ l of a lysis solution (50 mM KCl, 10 mM Tris pH 8.3, 2.5 mM MgCl₂, 0.01% gelatin, 0.2% SDS and 60 μ g ml^{–1} proteinase K) and incubated at 60 $^{\circ}$ C for 60 min with agitation. The samples were then frozen at –20 $^{\circ}$ C for later processing by HPLC. Each experiment was done in triplicate. A second method to measure DHP accumulation in worms gave qualitatively similar results (Supplementary Information).

See Supplementary Information for additional protocols. The pharmacological analysis of the egg-laying circuit was done as previously described^{10,24} and is presented in more detail in Supplementary Information. All primer sequences for mapping, archived pictures and worm strains are available on request.

Received 21 September 2005; accepted 14 February 2006.

- Mayer, T. U. *et al.* Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971–974 (1999).
- Loev, B., Ehrreich, S. J. & Tedeschi, R. E. Dihydropyridines with potent hypotensive activity prepared by the Hantzsch reaction. *J. Pharm. Pharmacol.* **24**, 917–918 (1972).
- Harrold, M. in *Foye's Principles of Medicinal Chemistry* (eds Williams, D. A. & Lemke, T. L.) 533–561 (Lippincott Williams and Wilkins, Baltimore, 2002).
- Lee, R. Y., Lobel, L., Hengartner, M., Horvitz, H. R. & Avery, L. Mutations in the α 1 subunit of an L-type voltage-activated Ca²⁺ channel cause myotonia in *Caenorhabditis elegans*. *EMBO J.* **16**, 6066–6076 (1997).
- Jospin, M., Jacquemond, V., Mariol, M. C., Segalat, L. & Allard, B. The L-type voltage-dependent Ca²⁺ channel EGL-19 controls body wall muscle function in *Caenorhabditis elegans*. *J. Cell Biol.* **159**, 337–348 (2002).
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
- Peterson, R. T., Link, B. A., Dowling, J. E. & Schreiber, S. L. Small molecule developmental screens reveal the logic and timing of vertebrate development. *Proc. Natl Acad. Sci. USA* **97**, 12965–12969 (2000).
- Priess, J. R. & Hirsh, D. I. *Caenorhabditis elegans* morphogenesis: the role of the cytoskeleton in elongation of the embryo. *Dev. Biol.* **117**, 156–173 (1986).

- Roy, P. J., Zheng, H., Warren, C. E. & Culotti, J. G. mab-20 encodes Semaphorin-2a and is required to prevent ectopic cell contacts during epidermal morphogenesis in *Caenorhabditis elegans*. *Development* **127**, 755–767 (2000).
- Trent, C., Tsuing, N. & Horvitz, H. R. Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* **104**, 619–647 (1983).
- Hockerman, G. H., Peterson, B. Z., Johnson, B. D. & Catterall, W. A. Molecular determinants of drug binding and action on L-type calcium channels. *Annu. Rev. Pharmacol. Toxicol.* **37**, 361–396 (1997).
- Steger, K. A., Shtonda, B. B., Thacker, C., Snutch, T. P. & Avery, L. The *C. elegans* T-type calcium channel CCA-1 boosts neuromuscular transmission. *J. Exp. Biol.* **208**, 2191–2203 (2005).
- Schafer, W. R. & Kenyon, C. J. A calcium-channel homologue required for adaptation to dopamine and serotonin in *Caenorhabditis elegans*. *Nature* **375**, 73–78 (1995).
- Franks, C. J. *et al.* Ionic basis of the resting membrane potential and action potential in the pharyngeal muscle of *Caenorhabditis elegans*. *J. Neurophysiol.* **87**, 954–961 (2002).
- Shtonda, B. & Avery, L. CCA-1, EGL-19 and EXP-2 currents shape action potentials in the *Caenorhabditis elegans* pharynx. *J. Exp. Biol.* **208**, 2177–2190 (2005).
- Stanley, E. F. & Atrakchi, A. H. Calcium currents recorded from a vertebrate presynaptic nerve terminal are resistant to the dihydropyridine nifedipine. *Proc. Natl Acad. Sci. USA* **87**, 9683–9687 (1990).
- Furukawa, T. *et al.* Selectivities of dihydropyridine derivatives in blocking Ca²⁺ channel subtypes expressed in *Xenopus* oocytes. *J. Pharmacol. Exp. Ther.* **291**, 464–473 (1999).
- Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the ventral nerve cord of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **275**, 327–348 (1976).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *C. elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Waggoner, L. E., Zhou, G. T., Schafer, W. R. & Schafer, W. R. Control of alternative behavioural states by serotonin in *Caenorhabditis elegans*. *Neuron* **21**, 203–214 (1998).
- Shyn, S. I., Kerr, R. & Schafer, W. R. Serotonin and G_o modulate functional states of neurons and muscles controlling *C. elegans* egg-laying behaviour. *Curr. Biol.* **13**, 1910–1915 (2003).
- Mathews, E. A. *et al.* Critical residues of the *Caenorhabditis elegans* unc-2 voltage-gated calcium channel that affect behavioural and physiological properties. *J. Neurosci.* **23**, 6537–6545 (2003).
- Schafer, W. R., Sanchez, B. M. & Kenyon, C. J. Genes affecting sensitivity to serotonin in *Caenorhabditis elegans*. *Genetics* **143**, 1219–1230 (1996).
- Dempsey, C. M., Mackenzie, S. M., Gargus, A., Blanco, G. & Sze, J. Y. Serotonin (5HT), fluoxetine, imipramine and dopamine target distinct 5HT receptor signaling to modulate *Caenorhabditis elegans* egg-laying behaviour. *Genetics* **169**, 1425–1436 (2005).
- Weinshenker, D., Wei, A., Salkoff, L. & Thomas, J. H. Block of an ether-a-go-go-like K⁺ channel by imipramine rescues *egl-2* excitation defects in *Caenorhabditis elegans*. *J. Neurosci.* **19**, 9831–9840 (1999).
- Lewis, J. A. & Fleming, J. T. in *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) (Academic, San Diego, 1995).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Turnbull, S. Ito and H. Cheung for technical assistance, S. Dixon, C. Boone and B. Andrews for advice on the manuscript, and A. Spence for sharing equipment. We thank the *C. elegans* Genetic Center, which is funded by the NIH National Center for Research Resources, for sending us worm strains. E.F.S., P.M., S.R.C. and P.J.R. are Canadian Research Chairs in brain and behaviour, plant molecular biology, plant genomics, and molecular neurobiology, respectively. This work was supported by an NSERC Industrial Grant to P.M., a CIHR Grant to E.F.S., a CIHR CGS to A.W.C., and a Premier's Research Excellence Award and awards from the Canadian Foundation for Innovation and Ontario Innovation Trust to P.J.R.

Author Contributions R.F. discovered that nemadipine-A induces the Vab phenotype, T.C.Y.K. and P.J.R. did the worm genetics and pharmacological analysis, P.M., R.F., A.B. and P.J.R. did the small-molecule screen, A.W.C. and E.F.S. did the electrophysiology, and S.R.C. did the HPLC analysis. N.R. provided technical assistance throughout. With critical input from S.R.C., P.J.R. led the project and wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests; details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.J.R. (peter.roy@utoronto.ca).

LETTERS

Downstream nuclear events in brassinosteroid signalling

Grégory Vert¹ & Joanne Chory^{1,2}

Brassinosteroids (BRs) are steroid hormones that control many aspects of plant growth and development^{1,2}. BRs bind to the plasma membrane receptor kinase BRI1, and act through a signalling pathway that involves a glycogen synthase kinase-3-like kinase (BIN2) and a serine/threonine phosphatase (BSU1)^{3–5}. Previous models proposed that BIN2 negatively regulates BR signalling by controlling the stability and subcellular localization of the related transcription factors BES1 and BZR1 by phosphorylation, in a manner reminiscent of the canonical Wnt signalling pathway of metazoans^{6–9}. Here we present strong evidence for a different mode of regulation of BR signalling. We show that BES1 is localized constitutively to the nucleus, where its activity is modulated by nuclear-localized BIN2 kinase. BIN2-mediated phosphorylation of BES1 inhibits its DNA-binding activity on BR-responsive target promoters and its transcriptional activity through impaired multimerization. Our observations demonstrate that phosphorylation-dependent inhibition of DNA binding and *trans*-activation is the key primary mechanism of BES1 regulation.

Polyhydroxylated steroid hormones have key functions in the growth and development of all multicellular eukaryotes. There is considerable conservation of steroid biosynthetic enzymes between plants and metazoans, yet there is no conservation of their signalling pathways. In animals, steroids are perceived by members of the nuclear receptor superfamily of ligand-dependent transcription factors^{10,11}. In contrast, plant steroids bind to a 100-amino-acid subdomain embedded within the extracellular leucine-rich repeats of the plasma membrane receptor kinase BRI1 (ref. 3).

Genetic defects in the synthesis or perception of BRs result in dark-green dwarf plants, with short stems, rounded leaves, delayed senescence, abnormal vascular development and reduced fertility. Knowledge of these phenotypes permitted the identification of gain-of-function mutations in *BIN2*, which encodes a glycogen synthase kinase-3 (GSK3)-type kinase, and indicates that this kinase might negatively regulate the pathway^{4,12,13}. Downstream from BIN2, two closely related transcription factors, BES1 and BZR1, bind BR target genes, either on their own or together with other transcription factors^{14,15}. BES1 and BZR1 are *in vitro* substrates of BIN2 and are phosphoproteins *in vivo*^{6,7,16}. When overexpressed as fusion proteins with green fluorescent protein (GFP) or cyan fluorescent protein, BES1 and BZR1 are rapidly dephosphorylated after treatment with BR^{6,7}, probably by the BSU1 phosphatase, a nuclear protein⁵. Additionally, it has been reported that these tagged versions of BES1 and BZR1 show slightly increased protein levels and nuclear localization upon BR application.

BR signalling was therefore proposed to follow a two-state model⁷, mechanistically similar to the canonical Wnt signalling pathway in metazoans, in which the phosphorylation by GSK3- β of β -catenin in the cytosol leads to its turnover in the absence of Wnt, whereas in the presence of Wnt, GSK3 β is inactivated, leading to the accumulation

and nuclear translocation of dephosphorylated β -catenin^{17,18}. Despite its attractiveness as a model for BR signalling there is little direct evidence for its existence, and there are conflicting observations about BES1 and BZR1 stabilization and subcellular localization¹⁹. In particular, the tagged versions of BES1 and BZR1 show slightly increased protein levels after treatment with BL in hypocotyl cells^{7,20}, which does not correlate with the dramatic shift in BES1 and BZR1 phosphorylation status. Further, the proposed model indicates that BES1 and BZR1 might be cytosolic proteins in the absence of BRs, but their cytosolic localization has not been documented and there are several lines of evidence that BES1 and BZR1 are constitutively nuclear^{16,19}. Last, BIN2, a candidate negative regulator of BR signalling, has been identified solely on the basis of its gain-of-function phenotype, and its function in *Arabidopsis* development, and in BR signalling in particular, remains unclear.

To clarify the downstream events in BR signalling, we identified a loss-of-function allele of *BIN2*, *bin2-3*, and showed that BIN2 acts in the pathway by promoting BES1 phosphorylation (Supplementary Fig. S1). In contrast to the typical blade-shaped leaves and long bending petioles observed in BL-treated or BL-hypersensitive plants overexpressing BRI1, for example²¹, *bin2-3* plants displayed no major visible phenotypes. This indicates that other GSK3s, among the ten found in the *Arabidopsis* genome, might act redundantly with BIN2 in the pathway. Although individual loss-of-function mutants of the two closest BIN2 homologues revealed no effect on BR signalling (Supplementary Fig. S2), gain-of-function mutations (E291K and E293K recapitulating the *bin2-1* mutation) indicated that these two GSK3s might have a function in BR signalling (Supplementary Fig. S3). Moreover, a triple knockout for group II GSK3s displayed constitutive BR responses and a strong resistance to the BR biosynthetic inhibitor brassinazole (BRZ; Fig. 1a–c). In the absence of treatment with BL the triple mutant had significantly higher levels of dephosphorylated BES1 than in wild-type plants (Fig. 1d), explaining its phenotype. The absence of total dephosphorylation of BES1 in the triple GSK3 mutant background indicates that a pool of BES1, probably not involved in hypocotyl or petiole elongation under such conditions, might not be under the control of these three GSK3s. This is supported by the wider expression pattern of BES1 previously reported⁷ than in the three GSK3s studied here (G. Vert and J. Chory, unpublished observations).

Having established that GSK3 kinases have a function in the BR signalling pathway, we next studied BIN2 regulation, because it represents a major signalling component in the BR response pathway and is the best-characterized GSK3 in plants. Previous reports indicate that the regulation of BIN2 might involve a mechanism that is different from the previously described mechanisms for the yeast and metazoan enzymes¹⁶. To gain insight into BIN2 regulation we generated transgenic plants expressing BIN2–GFP and *bin2*–GFP (containing the E263K gain-of-function mutation found in *bin2-1*) fusion proteins driven by the *BIN2* promoter. BIN2–GFP plants had a

¹Plant Biology Laboratory, and ²Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

wild-type phenotype and were slightly hypersensitive to BRZ (data not shown), whereas *bin2*-GFP plants were dwarves and recapitulated the *bin2-1* phenotype (Supplementary Fig. S4a). This indicates that both fusion proteins were functional *in vivo*. Neither dark/light growth conditions nor treatments with BRs or BRZ significantly affected BIN2 protein levels (Supplementary Fig. S4b). BIN2-GFP was found in both the nucleus and cytosol, as well as at the plasma membrane, although BIN2 has no obvious targeting sequences to either the nucleus or membranes (Fig. 2a). No change in the steady-state distribution of BIN2-GFP was observed in shifts from dark to light conditions or after treatments with BR or BRZ (data not shown). In contrast, the similar *bin2*-GFP fusion protein exhibited a strikingly different pattern of localization. When compared with lines expressing comparable levels of wild-type BIN2-GFP (Supplementary Fig. S5), *bin2*-GFP was found mostly localized to the nucleus (Fig. 2b).

These findings prompted us to reconsider the prevailing model of BR signalling in which BIN2 phosphorylates BES1 and BZR1 in the cytosol. To determine where in the cell BIN2 activity is required for the regulation of BR signalling, transgenic plants were constructed expressing mutant *bin2* protein fused to a heterologous nuclear localization signal (NLS), to a nuclear export signal (NES) or to non-functional forms of each (*nls* and *nes*, respectively). From more than 150 T₁ plants, a range of dwarf phenotypes was obtained; these were divided into five categories (Fig. 2c). When fused to an NLS, the distribution of dwarf plants was strongly shifted towards the most severe dwarf phenotypes compared with plants expressing *bin2* alone or *bin2* fused to the non-functional *nls*. In contrast, *bin2*-NES expression gave milder phenotypes than *bin2* alone or *bin2*-*nes*. These findings were confirmed with the wild-type form of BIN2 (Fig. 2c), and indicate that BIN2 is likely to act on BES1/BZR1 in the nucleus.

The finding that BIN2 is present and active in the nucleus, together with the previous observation that BSU1, a phosphatase that acts on phospho-BES1, is a nuclear protein⁵, prompted us to reinvestigate the proposed Wnt-based stabilization/translocation mechanism of BES1/BZR1 regulation by BIN2 in response to BRs. We monitored the endogenous BES1 protein by using an anti-BES1 antibody and confirmed that short-term treatment with BL resulted in a shift of BES1 to its dephosphorylated state, as previously reported for an overexpressed BES1-GFP fusion protein⁷ (Fig. 3a). However, no accumulation of BES1 was observed, even over a period of 4 h,

consistent with recent reports^{3,5}. This observation contrasts with the two- to three-fold changes that were reported in early studies^{6,7,20}. In addition, no major alteration in endogenous BES1 levels could be observed in BRZ-grown plants or in different genetic backgrounds affected in BR biosynthesis (*det2-1*) or signalling (*bri1-5* and *bin2-1*) (Fig. 3b). These conditions revealed an enhanced conversion of the BES1 pool to the phosphorylated form, which was consistent with their dwarf phenotype. Conversely, plants overexpressing BRI1 had higher levels of dephosphorylated BES1, probably explaining their hypersensitivity to BRs²¹. In BL-treated plants, BES1 was rapidly dephosphorylated in wild-type or *det2-1* plants, whereas the *bri1-5* and *bin2-1* insensitive mutants were largely unresponsive, as expected.

We next re-examined the BR-dependent change in localization of BES1 that was previously suggested, by looking at BES1-GFP under several conditions. We found that in both roots and hypocotyls BES1-GFP was readily detectable and constitutively nuclear (Fig. 3c). Using identical detection settings on the confocal microscope, we observed no difference in either the localization or levels of

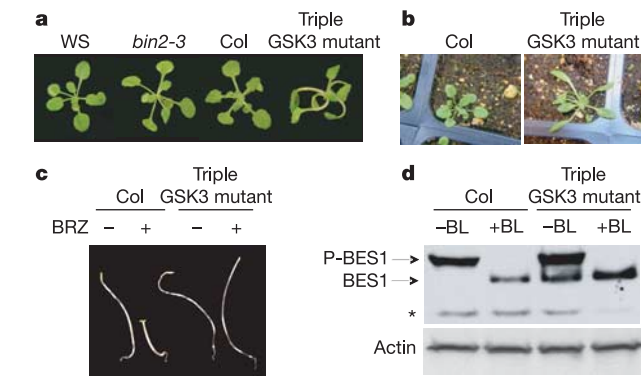


Figure 1 | GSK3 kinases act redundantly to negatively regulate BR signalling. **a–c**, Phenotype of the triple GSK3 mutant grown in long days and transferred to short-day conditions for 4 days (**a**), in soil (**b**), or in the presence of 1 μ M BRZ in the dark (**c**). **d**, BES1 phosphorylation status in the triple GSK3 mutant. P-BES1, phospho-BES1. Fifteen-day-old plants were submerged for 1 h in medium containing mock treatment or 1 μ M BL. Western blot analyses were performed with anti-BES1 antibody. Care was taken to load similar amounts of protein. The non-specific band indicated with an asterisk serves as a loading control. Actin antibody (ICN) was used as a second loading control.

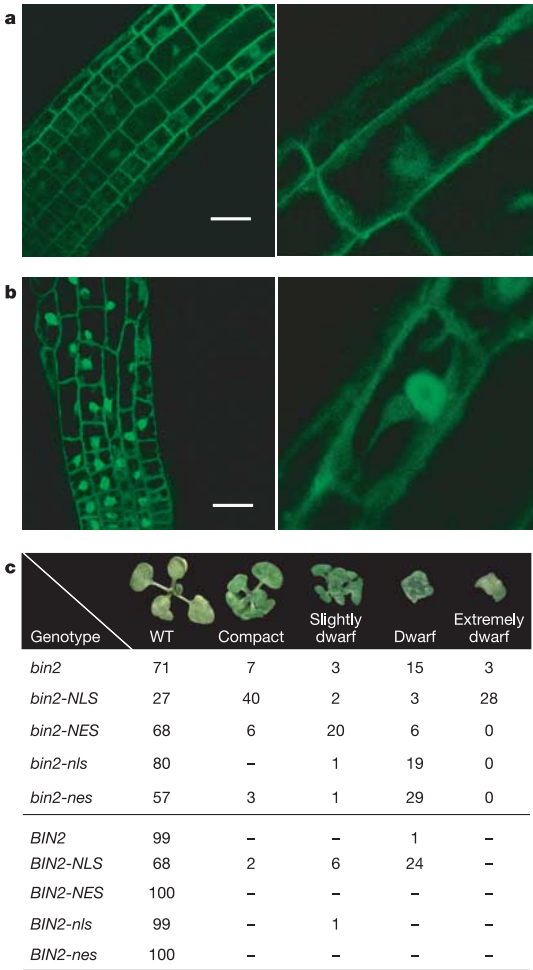


Figure 2 | BIN2 is found and required in the nucleus for BR signalling. **a**, BIN2 protein is found in the nucleus. Subcellular localization of BIN2-GFP by confocal laser scanning microscopy of line BIN2::BIN2-GFP no. 4 on root tips of light-grown seedlings. Root tip cells are poorly vacuolated and allow an easy visualization of both cytosolic and nuclear compartments. Scale bar, 20 μ m. **b**, Subcellular localization of *bin2*-GFP by confocal microscopy on line BIN2::*bin2*-GFP no. 1. Scale bar, 20 μ m. **c**, Phenotypes of plants expressing both wild-type BIN2 and the gain-of-function *bin2* protein form fused to the nuclear targeting sequences NLS, NES, *nls* and *nes*. Results are shown as percentage of phenotypes obtained according to the severity of dwarfism.

BES1–GFP in either tissue when plants were grown on 1 μ M BRZ. In addition, the same BES1–GFP line crossed to the *bin2-1* dwarf mutant showed comparable levels of fluorescence to the wild type and a clear nuclear localization pattern. Detailed time-course analyses and quantification of the fluorescent signal obtained after treatment with BL in either roots (Fig. 3d, movie in Supplementary Fig. S6a) or hypocotyls (Fig. 3e, movie in Supplementary Fig. S6b) revealed no significant effect on the distribution or the levels of BES1 protein, although such treatments alter the phosphorylation status of BES1. Taken together with our observations that BES1 protein does not accumulate to higher levels in response to BRs, these results indicate that neither protein stabilization nor nuclear translocation of BES1 is required for BR signalling, emphasizing the importance of phosphorylation in regulating BES1 activity.

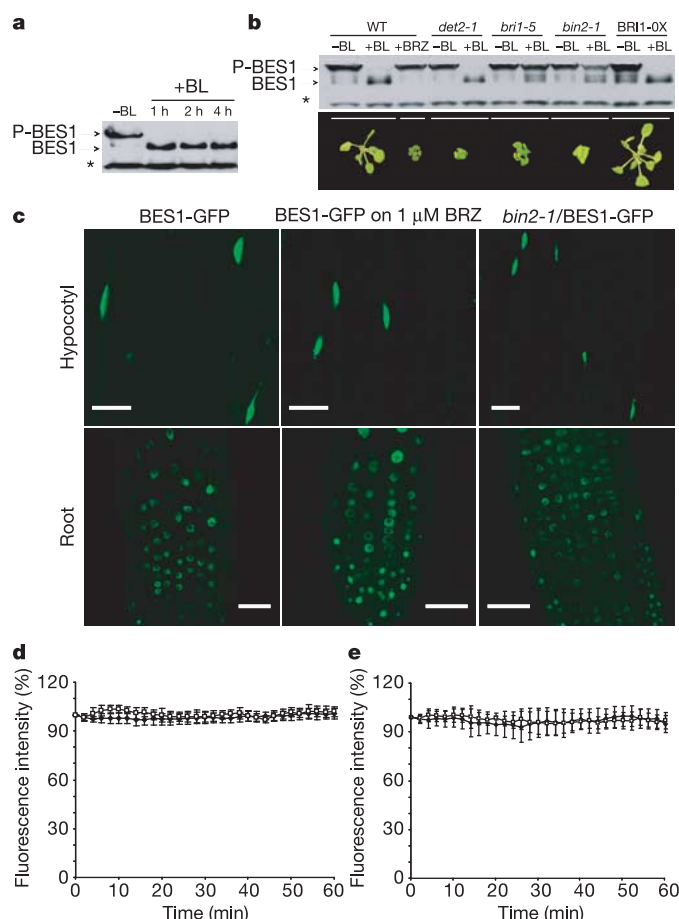


Figure 3 | BES1 is constitutively nuclear and its levels are unaffected by BRs. **a**, Time-course analyses of treatment with BL on BES1 protein. P-BES1, phospho-BES1. **b**, BES1 phosphorylation status in plants grown under different BR conditions (BL, BRZ), or in genotypes affected in BR biosynthesis (*det2-1*) or signalling (*bri1-5*, *bin2-1*, and a BRI1-OX plant overexpressing BRI1 (ref. 27)). Plants were treated as described in the legend to Fig. 1d and subjected to western blot analysis with anti-BES1 antibody. The phenotype of the corresponding plants is shown. Only the wild-type Col is shown because both Col and Ws ecotypes responded in the same way to treatment with BL (data not shown). **c**, Subcellular localization of BES1–GFP in root and hypocotyl cells. BES1 subcellular localization was investigated under normal growth conditions (left panels), on plants grown on 1 μ M BRZ (middle panels) and in the BR-insensitive *bin2-1* genetic background (right panels). Scale bars, 20 μ m. **d, e**, Quantification over time of BES1–GFP fluorescence levels after treatment with BL. Plants were subjected to either mock treatment (filled symbols) or 1 μ M BL (open symbols) followed by fluorescence analysis in roots (**d**) and hypocotyls (**e**) every 2 min for 1 h. Results are presented as means $\pm$ s.d. ($n = 30$) of individual nucleus fluorescence observed in five independent samples.

Having determined that the phosphorylation status of BES1 is the most reliable read-out of the BR signalling state, we examined how phospho-BES1 was different from its dephosphorylated form. Recently, BES1 and BZR1 have been shown to bind the 5'-CANNTG-3' E-box and 5'-CGTG(T/C)G-3' elements of BR-regulated genes, respectively^{14,15}. Using biotinylated DNA fragments, we confirmed the interaction between BES1 and both of these BR-responsive promoter elements (Fig. 4a). Treatment with calf intestinal phosphatase (CIP) had no effect on the binding of BES1 to DNA, probably because the protein is not phosphorylated in such an expression system. However, BIN2-mediated phosphorylation of BES1 resulted in a clear shift in its migration and in a complete loss of its DNA-binding activity, which was further confirmed by using BES1 recombinant protein with one of the promoters (Supplementary Fig. S7). We next investigated BES1 transcriptional activity in a heterologous yeast system, which assesses *trans*-activation independently of DNA-binding. A BES1–LexA fusion protein activated the transcription of a 6 $\times$ *LexAop::LEU2* gene in yeast and allowed growth on synthetic medium lacking leucine (Fig. 4b). However, co-expression with BIN2 almost completely alleviated transcriptional activation by BES1, although the yeast cells contained comparable levels of BES1–LexA fusion protein (Supplementary Fig. S8), indicating that BES1 phosphorylation might have abolished its transcriptional function.

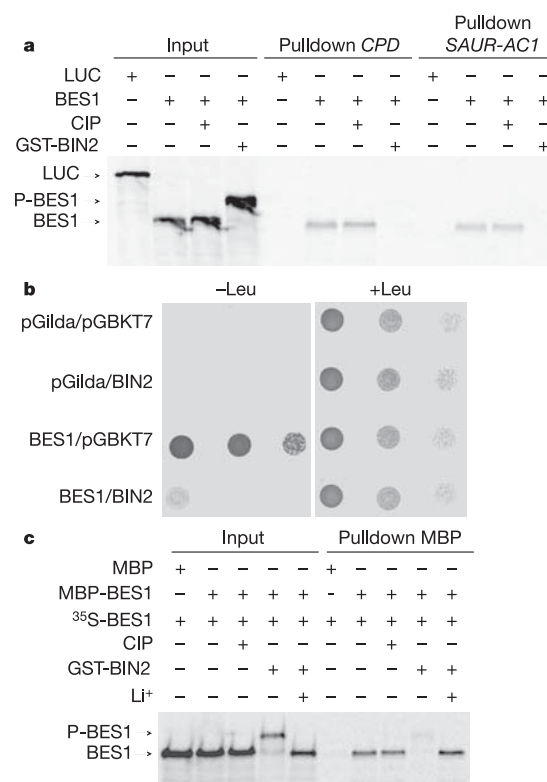


Figure 4 | BIN2-mediated phosphorylation abolishes BES1 DNA-binding and *trans*-activation. **a**, DNA–protein pull-down experiments with biotinylated DNA fragments from *CPD* and *SAUR-AC1* promoters with BES1 translated *in vitro*. BES1 protein was treated with CIP or incubated with BIN2 before pull-down experiments with DNA-bound streptavidin beads. Luciferase (LUC) is shown as negative control and inputs as loading control. P-BES1, phospho-BES1. **b**, BES1 transcriptional activity. Activation of the transcription of a *LexAop::LEU2* reporter by a BES1–LexA fusion protein was assessed by the growth in the absence of Leu. **c**, Protein–protein pull-down. BES1 translated *in vitro* was first subjected to phosphorylation modification by CIP, BIN2, or BIN2 inactivated with 100 mM Li⁺ before incubation with recombinant MBP or MBP–BES1 recombinant proteins. Pull-down was performed using amylose resin.

Previous studies have indicated that BES1 multimerizes for DNA binding¹⁵. We therefore tested whether the phosphorylation of BES1 influenced its multimerization. Dephosphorylated BES1 was indeed able to interact with itself *in vitro* (Fig. 4c), and its ability to multimerize was decreased by previous phosphorylation with BIN2. There was no effect on multimerization in the presence of the GSK3 inhibitor Li⁺, indicating that BIN2-mediated phosphorylation might directly block BES1 homo-multimerization.

These results demonstrate that the currently proposed model for BR signalling should be revised. Rather than regulating the nuclear translocation and accumulation of BES1, BIN2 seems to act in the nucleus to regulate the genomic response to BRs by phosphorylating BES1 and blocking both its binding to target promoters and its transcriptional activity (Supplementary Fig. S9). These results explain our previous *in vivo* observations in which we showed that more BES1 is bound to target promoters upon stimulation with BR¹⁵, which was originally interpreted to be due to increased nuclear translocation and stabilization of BES1.

Several lines of evidence indicate that BES1 and BZR1 proteins are degraded by the proteasome, including increased BES1 and BZR1 accumulation in the dominant *bes1-D* and *bzr1-D* mutants (which contain a proline to leucine change in a predicted PEST domain), or after treatments with MG132, a proteasome inhibitor^{6,7,20}. However, our data show that this regulation of protein levels is not a primary response to BRs, nor is it a requirement for BR signalling. Rather, BES1 and BZR1 are likely to be short-lived proteins, whose total levels (but not the ratio of phosphorylated to dephosphorylated forms) may be altered by environmental growth conditions (Y. Tao and J.C., unpublished data). Microarray analyses have revealed that *BZR1*, but not *BES1*, is slightly induced by treatment with BR¹⁹, which might explain the mild accumulation of a BZR1 fusion protein after treatments with BR²⁰. However, this increase is unlikely to account for the phenotypic effects observed. Nevertheless, a careful re-examination of native BES1 or BES1-GFP levels after treatment with BR indicates that BES1 levels might not change in response to the hormone.

Regulated turnover or accumulation of key nuclear proteins has emerged as a common theme in most plant hormone signalling pathways, including ethylene²², abscisic acid²³, gibberellins²⁴, auxin²⁵ and probably jasmonate²⁶. In contrast, BR perception leads to a phosphorylation cascade that culminates in the nuclear dephosphorylation of BES1. This leads to multimerization of BES1 with itself or other transcription factors, which allows it to bind and *trans*-activate the expression of scores of target genes⁷. As such, identifying BES1 target genes and understanding how these gene expression changes lead to increases in growth and plant stature may now be possible.

METHODS

DNA binding assays. GST-BIN2 (Amersham Pharmacia) and maltose-binding protein (MBP)-BES1 (New England Biolabs) were prepared in accordance with the manufacturer's recommendations. In parallel, BES1 was subjected to [³⁵S]Met TNT T7 *in vitro* transcription/translation with a reticulocyte lysate system (Promega). Both *in vitro* translated and recombinant BES1 proteins were treated with CIP or BIN2 as described¹⁶. Biotinylated DNA fragments corresponding to *CPD* and *SAUR-AC1* promoters were generated by polymerase chain reaction, as described^{14,15}. For DNA-protein pull-down, biotinylated promoters were first incubated with streptavidin-bound agarose beads (Sigma) in IP100 buffer (100 mM potassium glutamate, 50 mM Tris-HCl pH 7.6, 2 mM MgCl₂, 0.05% Nonidet P40) for 2 h at 4 °C, then washed three times in IP100 buffer. Proteins were added to DNA-bound beads and the mixture was rotated in a 4 °C cold room for 2 h. Beads were washed three times with IP100 buffer; proteins were stripped off the beads by boiling with 2 × SDS buffer and then subjected to PAGE. The gel was fixed and dried; signals were detected with a PhosphorImager (Molecular Dynamics).

Transcription activation. *BES1* was cloned in pGilda to create a BES1-LexA fusion protein. *BIN2* was cloned in pGBKT7. Yeast (strain EGY48), expressing the *LEU2* gene under the control of 6 × *LexA* operators, was co-transformed with the above constructs or the corresponding empty vectors, and transcription

activation was assayed by growth on galactose-containing medium lacking leucine.

Protein-protein interaction assays. BES1 translated *in vitro* was incubated with CIP and BIN2 in the presence or absence of 100 mM lithium chloride as described previously¹⁶. The reaction mixtures were then incubated in IP100 buffer with recombinant MBP or MBP-BES1 proteins for 2 h in the cold room. Amylose resin was then added to the reaction mixture and incubated for 2 h at 4 °C. Subsequent steps were performed as described above.

Confocal microscopy. Confocal microscopy analyses were performed with a Leica SP2 confocal microscope. For fluorescence quantification analyses, samples were embedded on the microscope slide in LS/2 medium containing mock treatment or 1 μM BL. The fluorescence of 5-μm stacks, encompassing a complete cell layer of the sample to avoid the movement of nuclei in the cell, was recorded every 2 min over a period of 1 h. Images were then subjected to maximum orthogonal projection (Leica software) over the whole stack, and aligned to correct for growth and movement with ImageJ software. The fluorescence of individual nuclei was monitored and quantified with ImageJ software over the time course of the experiment.

Received 30 January; accepted 28 February 2006.

- Clouse, S. D., Langford, M. & McMorris, T. C. A brassinosteroid-insensitive mutant in *Arabidopsis thaliana* exhibits multiple defects in growth and development. *Plant Physiol.* **111**, 671–678 (1996).
- Mandava, N. B. Plant growth-promoting brassinosteroids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**, 23–52 (1988).
- Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005).
- Li, J. & Nam, K. H. Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. *Science* **295**, 1299–1301 (2002).
- Mora-Garcia, S. *et al.* Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in *Arabidopsis*. *Genes Dev.* **18**, 448–460 (2004).
- He, J. X., Gendron, J. M., Yang, Y., Li, J. & Wang, Z. Y. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 10185–10190 (2002).
- Yin, Y. *et al.* BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **109**, 181–191 (2002).
- Clouse, S. D. Brassinosteroid signal transduction: Clarifying the pathway from ligand perception to gene expression. *Mol. Cell* **10**, 973–982 (2002).
- Peng, P. & Li, J. Brassinosteroid signal transduction: A mix of conservation and novelty. *J. Plant Growth Regul.* **22**, 298–312 (2003).
- Hall, J. M., Couse, J. F. & Korach, K. S. The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J. Biol. Chem.* **276**, 36869–36872 (2001).
- Olefsky, J. M. Nuclear receptor minireview series. *J. Biol. Chem.* **276**, 36863–36864 (2001).
- Choe, S. *et al.* *Arabidopsis* brassinosteroid-insensitive *dwarf12* mutants are semidominant and defective in a glycogen synthase kinase 3β-like kinase. *Plant Physiol.* **130**, 1506–1515 (2002).
- Perez-Perez, J. M., Ponce, M. R. & Micol, J. L. The *UCU1 Arabidopsis* gene encodes a SHAGGY/GSK3-like kinase required for cell expansion along the proximodistal axis. *Dev. Biol.* **242**, 161–173 (2002).
- He, J. X. *et al.* BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and growth responses. *Science* **307**, 1634–1638 (2005).
- Yin, Y. *et al.* A new class of transcription factors mediates brassinosteroid-regulated gene expression in *Arabidopsis*. *Cell* **120**, 249–259 (2005).
- Zhao, J. *et al.* Two putative BIN2 substrates are nuclear components of brassinosteroid signaling. *Plant Physiol.* **130**, 1221–1229 (2002).
- Salic, A., Lee, E., Mayer, L. & Kirschner, M. W. Control of β-catenin stability: Reconstitution of the cytoplasmic steps of the wnt pathway in *Xenopus* egg extracts. *Mol. Cell* **5**, 523–532 (2000).
- Staal, F. J., Noort Mv, M., Strous, G. J. & Clevers, H. C. Wnt signals are transmitted through N-terminally dephosphorylated β-catenin. *EMBO Rep.* **3**, 63–68 (2002).
- Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. & Chory, J. Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21**, 177–201 (2005).
- Wang, Z. Y. *et al.* Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. *Dev. Cell* **2**, 505–513 (2002).
- Wang, Z. Y., Seto, H., Fujioka, S., Yoshida, S. & Chory, J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature* **410**, 380–383 (2001).
- Guo, H. & Ecker, J. R. Plant responses to ethylene gas are mediated by SCF(EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. *Cell* **115**, 667–677 (2003).

23. Lopez-Molina, L., Mongrand, S. & Chua, N. H. A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **98**, 4782–4787 (2001).
24. Sasaki, A. *et al.* Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. *Science* **299**, 1896–1898 (2003).
25. Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. *Nature* **414**, 271–276 (2001).
26. Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. COI1: An *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280**, 1091–1094 (1998).
27. Friedrichsen, D. M., Joazeiro, C. A., Li, J., Hunter, T. & Chory, J. Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiol.* **123**, 1247–1256 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Nemhauser, N. Geldner and A. Nott for critical reading of the manuscript and discussions; T. Dabi and S. Palida for technical assistance; K. Chen for sharing the vectors containing NLS, nls, NES and nes; and Y. Yin for providing the BES1 antibody. This work was supported by a Human Frontier Science Program Organization long-term fellowship to G.V. and grants to J.C. from the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, and the National Science Foundation. J.C. is an investigator of the Howard Hughes Medical Institute.

Author Contributions G.V. and J.C. conceived and designed the experiments. G.V. performed the experiments. G.V. and J.C. analyzed the data and wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.C. (chory@salk.edu).

Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses

Hiroki Kato^{1,3*}, Osamu Takeuchi^{1,3*}, Shintaro Sato³, Mitsutoshi Yoneyama⁴, Masahiro Yamamoto¹, Kosuke Matsui¹, Satoshi Uematsu¹, Andreas Jung¹, Taro Kawai³, Ken J. Ishii³, Osamu Yamaguchi⁵, Kinya Otsu⁵, Tohru Tsujimura⁶, Chang-Sung Koh⁷, Caetano Reis e Sousa⁸, Yoshiharu Matsuura², Takashi Fujita⁴ & Shizuo Akira^{1,3}

The innate immune system senses viral infection by recognizing a variety of viral components (including double-stranded (ds)RNA) and triggers antiviral responses^{1,2}. The cytoplasmic helicase proteins RIG-I (retinoic-acid-inducible protein 1, also known as Ddx58) and MDA5 (melanoma-differentiation-associated gene 5, also known as Ifih1 or Helicard) have been implicated in viral dsRNA recognition^{3–7}. *In vitro* studies suggest that both RIG-I and MDA5 detect RNA viruses and polyinosine-polycytidylic acid (poly(I:C)), a synthetic dsRNA analogue³. Although a critical role for RIG-I in the recognition of several RNA viruses has been clarified⁸, the functional role of MDA5 and the relationship between these dsRNA detectors *in vivo* are yet to be determined. Here we use mice deficient in MDA5 (*MDA5*^{−/−}) to show that MDA5 and RIG-I recognize different types of dsRNAs: MDA5 recognizes poly(I:C), and RIG-I detects *in vitro* transcribed dsRNAs. RNA viruses are also differentially recognized by RIG-I and MDA5. We find that RIG-I is essential for the production of interferons in response to RNA viruses including paramyxoviruses, influenza virus and Japanese encephalitis virus, whereas MDA5 is critical for picornavirus detection. Furthermore, *RIG-I*^{−/−} and *MDA5*^{−/−} mice are highly susceptible to infection with these respective RNA viruses compared to control mice. Together, our data show that RIG-I and MDA5 distinguish different RNA viruses and are critical for host antiviral responses.

Host pattern recognition receptors, such as Toll-like receptors (TLRs) and helicase family members, have an essential role in the recognition of molecular patterns specific for different viruses, including DNA, single-stranded (ss)RNA, dsRNA and glycoproteins^{1,9,10}. dsRNA can be generated during viral infection as a replication intermediate for RNA viruses. TLR3, which localizes in the endosomal membrane, has been shown to recognize viral dsRNA as well as the synthetic dsRNA analogue poly(I:C) (refs 11, 12). The cytoplasmic proteins RIG-I and MDA5 have also been identified as dsRNA detectors^{3–5,7,13}. RIG-I and MDA5 contain two caspase-recruitment domains (CARDs) and a DExD/H-box helicase domain. RIG-I recruits a CARD-containing adaptor, IPS-1 (also known as MAVS, VISA or Cardif)^{14–17}. IPS-1 relays the signal to the kinases TBK1 and IKK-i, which phosphorylate interferon-regulatory factor-3 (IRF-3) and IRF-7, transcription factors essential for the expression of type-I

interferons^{18–22}. In contrast, TLR3 activates TBK1 and IKK-i through the TIR-domain-containing adaptor TRIF (also known as Ticam1)¹².

In vitro studies have shown that both RIG-I and MDA5 can bind to poly(I:C) and respond to poly(I:C) and RNA viruses⁶. We have generated *RIG-I*^{−/−} mice, and show that RIG-I is essential eliciting the immune responses against several RNA viruses, including Newcastle disease virus (NDV), Sendai virus (SeV) and vesicular stomatitis virus (VSV), in various cells except for plasmacytoid dendritic cells (pDCs)⁸. Hepatitis C virus and Japanese encephalitis virus are also reported to be recognized by RIG-I *in vitro*^{23,24}.

The *in vivo* functional relationship between RIG-I and MDA5 remains to be determined. To investigate a functional role for MDA5 *in vivo*, we generated *MDA5*^{−/−} mice and investigated viral recognition (Supplementary Fig. 1). In contrast to *RIG-I*^{−/−} mice, which are mostly embryonic lethal, *MDA5*^{−/−} mice are born in a mendelian ratio, grow healthily and do not show gross developmental abnormalities until 24 weeks of age. Flow cytometric analysis of leukocytes from the spleen and lymph nodes (staining for CD3, B220 and CD11c) revealed that the composition of lymphocytes and dendritic cells is similar in wild-type and *MDA5*^{−/−} mice (data not shown).

TLR3, RIG-I and MDA5 have been implicated in the recognition of poly(I:C) and the subsequent induction of antiviral responses. However, their exact contribution to *in vivo* responses against dsRNA has yet to be clarified. We therefore examined the *in vivo* responses to poly(I:C) in mice lacking RIG-I, MDA5 or TRIF, or both MDA5 and TRIF. Administration of poly(I:C) led to rapid induction of the cytokines interferon-α (IFN-α), IFN-β, interleukin-6 (IL-6) and IL-12 in sera of both wild-type and *RIG-I*^{−/−} mice (Fig. 1a and Supplementary Fig. 2a). In contrast, *MDA5*^{−/−} mice failed to produce IFN-α and IFN-β in response to poly(I:C), and production of IL-6 and IL-12p40 was also significantly impaired (Fig. 1b). Although *Trif*^{−/−} mice produced normal amounts of IFN-α, they also showed severely impaired production of IL-12p40 and partial impairment in IL-6 production. *MDA5*^{−/−}; *Trif*^{−/−} double-knock-out mice failed to induce IFN-α, IL-6 and IL-12p40 in response to poly(I:C). These results indicate that MDA5 is essential for poly(I:C)-induced IFN-α production and TLR3 signalling is critical for IL-12 production, whereas both MDA5 and TLR3 regulate IL-6 production.

¹Department of Host Defense, ²Department of Molecular Virology, Research Institute for Microbial Diseases, Osaka University, and ³ERATO, Japan Science and Technology Agency, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁴Department of Genetics and Molecular Biology, Institute for Virus Research, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. ⁵Department of Cardiovascular Medicine, Osaka University Graduate School of Medicine, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁶Department of Pathology, Hyogo College of Medicine, 1-1 Mukogawa-cho, Nishinomiya, Hyogo 663-8501, Japan. ⁷Department of Medical Technology, Shinshu University School of Allied Medical Sciences, 3-1-1 Asahi, Matsumoto 390-8621, Japan. ⁸Immunobiology Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.

*These authors contributed equally to this work.

When bone-marrow-derived dendritic cells generated by granulocyte-macrophage colony-stimulating factor (GM-CSF) were incubated in the presence of poly(I:C), production of IFN- α and IFN- β was severely impaired in *MDA5*^{-/-}, but not in *RIG-I*^{-/-} or *Trif*^{-/-}, GM-CSF-DCs (Fig. 1c and Supplementary Fig. 2b). Even when poly(I:C) was transfected into GM-CSF-DCs using lipofectamine, poly(I:C) induced IFN- β production in an MDA5-dependent, but not a RIG-I- or TRIF-dependent, manner (Fig. 1d). IFN- β production in response to poly(I:C) was also impaired in *MDA5*^{-/-} mouse embryonic fibroblasts (MEFs) (Fig. 1e), indicating that poly(I:C) is primarily recognized by MDA5, not RIG-I and TLR3, in these cells.

dsRNAs transcribed *in vitro* (Supplementary Fig. 2c) also stimulated MEFs to produce IFN- β . Unlike for poly(I:C), wild-type and *MDA5*^{-/-} MEFs produced comparable amounts of IFN- β (Fig. 1e) in response to *in vitro* transcribed dsRNAs. In contrast, *RIG-I*^{-/-} MEFs did not produce detectable amounts of IFN- β , indicating that RIG-I is essential for the detection of *in vitro* transcribed dsRNAs. As RIG-I, but not MDA5, is responsible for IFN- β production in response to dsRNAs of various lengths, these helicases probably distinguish nucleotide structure or sequence, but not length. Together, these results indicate that MDA5 and RIG-I are involved

in the detection of poly(I:C) and *in vitro* transcribed dsRNAs, respectively.

This finding led us to hypothesize that RIG-I and MDA5 are involved in the detection of different RNA viruses. We have previously shown that a set of negative-sense RNA viruses are recognized by RIG-I⁸. We first examined IFN- β and IFN- α production in *MDA5*^{-/-} MEFs in response to a set of negative-sense ssRNA viruses, including NDV, SeV, VSV and influenza virus. As infection with most of the wild-type viruses (except NDV) failed to induce type-I interferons in MEFs, owing to suppression of interferon responses by viral proteins (data not shown), we also used mutant viruses lacking viral interferon-inhibitory proteins. As shown in Fig. 2a and Supplementary Fig. 4b, wild-type MEFs produce IFN- β and IFN- α in response to these mutant viruses. Production of type-I interferons was severely impaired in *RIG-I*^{-/-} MEFs compared to wild-type cells, but MDA5 was dispensable for the production of type-I interferons. Japanese encephalitis virus (JEV), a positive-sense ssRNA virus belonging to the flavivirus family, also required RIG-I, but not MDA5, for IFN- β production (Fig. 2b).

We then examined the interferon responses of MEFs to encephalomyocarditis virus (EMCV), a positive-sense ssRNA virus belonging to the picornavirus family. EMCV-induced IFN- β production was abrogated in *MDA5*^{-/-} MEFs (Fig. 2c). In contrast, wild-type and *RIG-I*^{-/-} MEFs produced comparable amounts of IFN- β , indicating that EMCV is specifically recognized by MDA5. The induction of genes encoding IFN- β , IP-10 and IL-6 in response to EMCV was abrogated in *MDA5*^{-/-} macrophages (Supplementary Fig. 3d). The synthesis of cellular proteins in *MDA5*^{-/-} MEFs was progressively inhibited during EMCV infection, to an extent and with kinetics similar to wild-type MEFs (Supplementary Fig. 5), indicating that the EMCV infection was established in wild-type and *MDA5*^{-/-} MEFs in a similar manner. Moreover, other viruses belonging to the picornavirus family (Theiler's and Mengo viruses) also induced IFN- α through MDA5 (Supplementary Fig. 4d). Furthermore, the production of IFN- β in response to SeV and EMCV was impaired in *RIG-I*^{-/-} and *MDA5*^{-/-} GM-CSF-DCs, respectively (Fig. 2d, e), indicating that conventional dendritic cells (cDCs) also use these helicases for the differential recognition of viruses. EMCV-induced production of IL-6 was also abrogated in *MDA5*^{-/-}, but not *RIG-I*^{-/-}, cDCs (Supplementary Fig. 4c). Therefore, MDA5 is critical for the regulation of pro-inflammatory cytokines as well as type-I interferons in response to EMCV.

We next examined whether viral RNAs derived from VSV and EMCV recapitulate the production of interferons through MDA5 and RIG-I. When transfected into GM-CSF-DCs by lipofection, RNAs prepared from VSV or EMCV induced production of IFN- α in a RIG-I- or MDA5-dependent manner, respectively (Fig. 2f). We also performed reconstitution experiments by transfecting RIG-I or MDA5 expression vectors into *RIG-I*^{-/-}; *MDA5*^{-/-} MEFs, in which IFN- β induction was completely abrogated in response to infection with EMCV or SeV Cm (SeV with a mutated C protein) (Fig. 2g). The ectopic expression of human RIG-I, but not MDA5, activated the *Ifnb* promoter in response to SeV Cm. Reciprocally, cells expressing human MDA5, but not RIG-I, activated the *Ifnb* promoter in response to EMCV in a dose-dependent manner (Fig. 2h). These results indicate that human RIG-I and MDA5 recognize different RNA viruses by recognizing viral RNAs.

Previous studies have shown that pDCs use mainly the TLR system instead of RIG-I in the recognition of several RNA viruses⁸. MyD88 is an adaptor protein essential for TLR signalling (except through TLR3). We purified B220⁺ pDCs from Flt3L-generated bone-marrow-derived dendritic cells (Flt3L-DCs) and infected them with EMCV. pDCs from *Myd88*^{-/-}, but not *MDA5*^{-/-}, mice showed a profound defect in IFN- α production (Supplementary Fig. 6). Reciprocally, MDA5, but not MyD88, is required for the production of IFN- α in B220⁺ cDCs purified from Flt3L-DCs (Supplementary Fig. 6). These results indicate that both MDA5 and RIG-I are

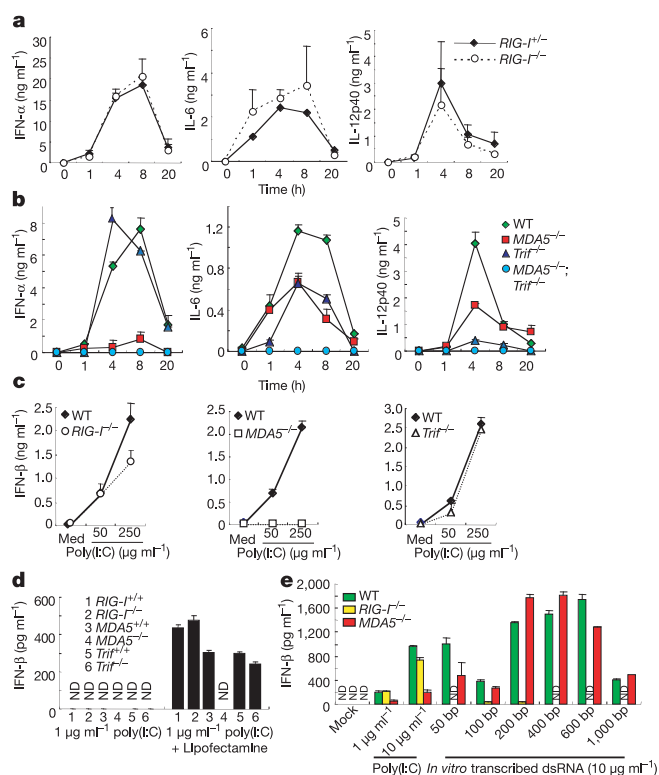


Figure 1 | Roles of MDA5, RIG-I and TRIF in the recognition of synthesized dsRNAs and dsRNA analogues. **a**, *RIG-I*^{-/-} and littermate *RIG-I*^{+/+} mice (a) or wild-type (WT), *MDA5*^{-/-}, *Trif*^{-/-} or *MDA5*^{-/-}; *Trif*^{-/-} double-knockout mice (**b**) were injected intravenously with 200 μ g poly(I:C) for the indicated periods, and IFN- α , IL-6 and IL-12p40 production was measured in serum by ELISA. Data show mean $\pm$ s.d. **c**, GM-CSF-DCs from *RIG-I*^{-/-}, *MDA5*^{-/-}, *TRIF*^{-/-} and littermate control mice were incubated in the presence of 50 or 250 μ g ml⁻¹ poly(I:C) for 24 h. IFN- β production in the cell culture supernatants was measured by ELISA. Med, medium only. **d**, GM-CSF-DCs were treated with 1 μ g ml⁻¹ poly(I:C) complexed with or without lipofectamine 2000 for 24 h, and IFN- β production was measured. **e**, Wild-type, *RIG-I*^{-/-} and *MDA5*^{-/-} MEFs were treated with poly(I:C) or *in vitro* transcribed dsRNAs of indicated lengths complexed with lipofectamine 2000 for 12 h, and IFN- β production was measured. Error bars indicate s.d. of triplicate wells in a single experiment; data are representative of three independent experiments. ND, not detected.

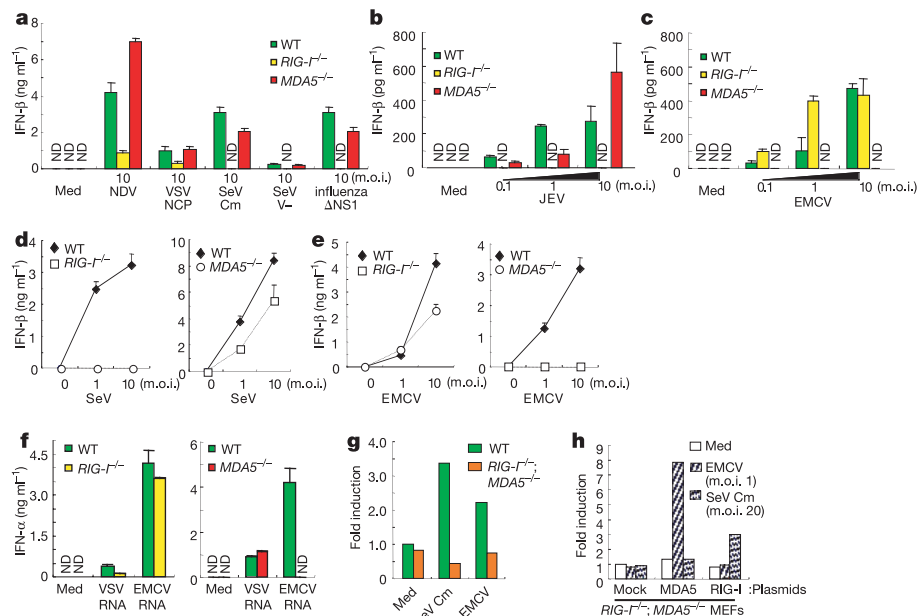


Figure 2 | Differential viral recognition by RIG-I and MDA5. **a**, Wild-type, *RIG-I*^{-/-} and *MDA5*^{-/-} MEFs were exposed to negative-sense ssRNA viruses, including NDV, VSV lacking a variant of M protein (NCP), SeV with a mutated C protein (Cm), SeV lacking V protein (V⁻), and influenza virus lacking the NS1 protein (ΔNS1) for 24 h. IFN-β production in the culture supernatants was measured by ELISA. **b**, **c**, Wild-type, *RIG-I*^{-/-} and *MDA5*^{-/-} MEFs were exposed to the positive-sense ssRNA viruses JEV (**b**) and EMCV (**c**), and IFN-β production was measured. **d**, **e**, GMSCF-DCs from *RIG-I*^{-/-} and *MDA5*^{-/-} mice and their littermate wild-type mice were infected with an increasing m.o.i. of SeV V⁻ (**d**) or EMCV (**e**) for 24 h, and IFN-β production was measured. **f**, Wild-type, *RIG-I*^{-/-} and *MDA5*^{-/-}

GMSCF-DCs were treated with RNAs directly prepared from VSV and EMCV (complexed with lipofectamine 2000) for 24 h, and IFN-α production was measured. **g**, Wild-type and *RIG-I*^{-/-}; *MDA5*^{-/-} MEFs were transiently transfected with a reporter construct containing the *Ifnb* promoter and exposed to SeV Cm or EMCV for 24 h. Cell lysates were then prepared and subjected to a luciferase assay. **h**, *RIG-I*^{-/-}; *MDA5*^{-/-} MEFs were transiently transfected with the *Ifnb* promoter construct together with expression plasmids encoding human RIG-I or MDA5. The cells were then infected with EMCV or SeV Cm for 24 h and were subjected to a luciferase assay. Error bars in **a–g** indicate s.d. of triplicate wells in a single experiment; data are representative of three independent experiments. ND, not detected.

dispensable for the viral induction of IFN-α in pDCs.

We next examined the *in vivo* roles of MDA5 and RIG-I in host defence against viral infection. Although most *RIG-I*^{-/-} mice are embryonic lethal⁸, we could efficiently obtain live adult mice by intercrossing the *RIG-I*^{+/-} mice obtained after *RIG-I*^{+/-} × ICR crosses (Supplementary Table 1). When the mice were infected with JEV, serum IFN-α levels were markedly decreased in *RIG-I*^{-/-} mice compared to littermate *RIG-I*^{+/-} mice. In contrast, *MDA5*^{-/-} mice did not show a defect in JEV-induced systemic IFN-α production (Fig. 3a). IFN-α production was partially impaired in *Myd88*^{-/-} mice compared to wild-type mice, but the extent of this impairment was far less than in *RIG-I*^{-/-} mice (Fig. 3a). These data suggest that the TLR system is not critical for the induction of serum IFN-α *in vivo* in response to JEV. Consistent with this finding, *RIG-I*^{-/-} mice, but not *MDA5*^{-/-} or *Myd88*^{-/-} mice, were more susceptible to JEV infection than control mice (Fig. 3b). Furthermore, *RIG-I*^{-/-} mice, but not *MDA5*^{-/-} mice, succumbed to VSV infection, consistent with abrogated interferon responses (Supplementary Fig. 7). Thus, RIG-I-mediated recognition of a specific set of viruses has a critical role in antiviral host defence *in vivo*.

We next challenged the mice with EMCV as a model virus that is recognized by MDA5. Induction of IFN-β, IFN-α, RANTES and IL-6 was severely impaired in the sera of *MDA5*^{-/-} mice (Fig. 4a and Supplementary Fig. 8). *MDA5*^{-/-} mice and mice null for the IFN-α/β receptor (*Ifnar1*^{-/-}) were highly susceptible to EMCV infection (viral titre of 1 × 10² plaque-forming units (p.f.u.) compared to littermate controls (*P* < 0.01) (Fig. 4b). In contrast, deficiency of neither RIG-I nor TLR3 affected the survival of mice infected with EMCV. Consistent with a previous report²², *Myd88*^{-/-} mice were modestly susceptible to EMCV infection compared to wild-type mice, implying that pDC-mediated responses are not critical for eliminating EMCV (Fig. 4b).

It is known that EMCV preferentially infects cardiomyocytes and causes myocarditis. Consistent with increased susceptibility to EMCV, viral titre in the heart was much higher in *MDA5*^{-/-} mice compared to control mice (Fig. 4c). Histological analysis of hearts two days after EMCV infection revealed that focal necrosis of

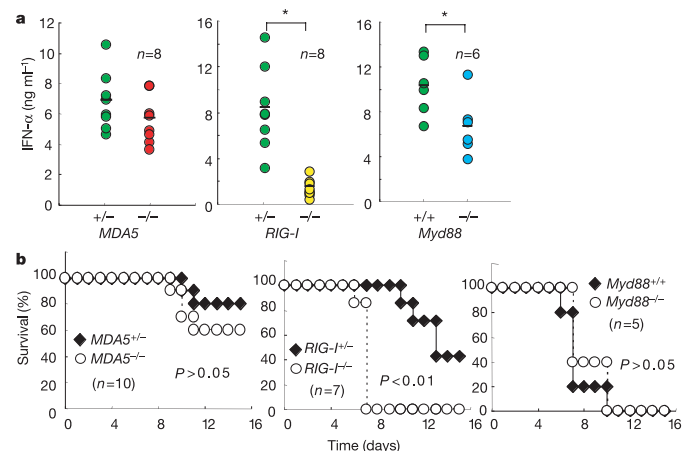


Figure 3 | Susceptibility of *RIG-I*^{-/-} and *MDA5*^{-/-} mice to JEV infection. **a**, *RIG-I*^{+/-}, *RIG-I*^{-/-}, *MDA5*^{+/-} and *MDA5*^{-/-} mice (*n* = 8), and *Myd88*^{+/+} or *Myd88*^{-/-} mice (*n* = 6), were injected intravenously with 2 × 10⁷ p.f.u. JEV. Sera were collected 24 h after injection, and IFN-α production levels measured by ELISA. Circles represent individual mice, bars indicate mean values. Asterisk, *P* < 0.05 versus controls (*t*-test). **b**, The survival of 6-week-old mice (genotypes as indicated) infected intravenously with 2 × 10⁷ p.f.u. JEV. Mice were monitored for 15 days (*P* < 0.01 between *RIG-I*^{-/-} mice and their littermate controls, generalized Wilcoxon test).

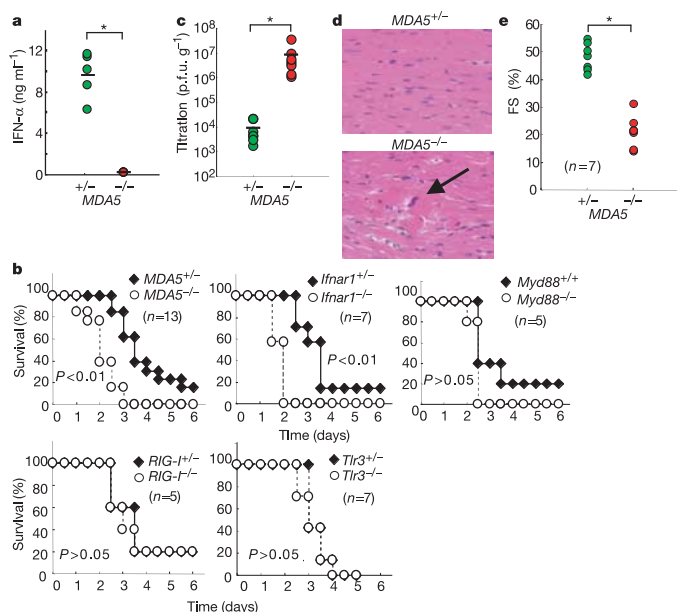


Figure 4 | Role of MDA5 in host defence against EMCV infection.

a, $MDA5^{+/+}$ and $MDA5^{-/-}$ mice ($n = 5$) were inoculated intravenously with 1×10^7 p.f.u. EMCV. Sera were prepared 4 h after injection and IFN- α production levels determined by ELISA. **b**, The survival of 6-week-old mice (genotypes as indicated) infected with 1×10^2 p.f.u. EMCV intraperitoneally was monitored every 12 h for six days ($P < 0.01$ between $MDA5^{-/-}$ or $Ifnar1^{-/-}$ mice and their littermate controls, generalized Wilcoxon test). **c**, $MDA5^{+/+}$ and $MDA5^{-/-}$ mice were infected intraperitoneally with 1×10^2 p.f.u. EMCV. After 48 h, mice were killed and virus titres in hearts were determined by plaque assay. **d**, Heart sections of $MDA5^{+/+}$ and $MDA5^{-/-}$ mice, two days after infection, were assessed for histological changes using haematoxylin and eosin staining. Arrow indicates the focal necrosis of cardiomyocytes. **e**, Cardiac function of mice 48 h after EMCV infection was assessed by echocardiography (see Supplementary Fig. 8b). The fractional shortening (FS) after infection determined by transthoracic M-mode echocardiographic tracings is shown. Asterisk, $P < 0.05$ versus $MDA5^{+/+}$ mice (t -test).

cardiomyocytes had developed in $MDA5^{-/-}$ mice, but wild-type hearts showed no histological abnormalities at this time point (Fig. 4d). Notably, no infiltration of immune cells was observed in either wild-type or $MDA5^{-/-}$ heart sections at this time point. However, when cardiac performance was analysed by echocardiography two days after infection (Fig. 4e), cardiac contractility was severely depressed in $MDA5^{-/-}$ mice (fractional shortening $48.2 \pm 4.9\%$ in $MDA5^{+/+}$ mice, $21.2 \pm 5.8\%$ in $MDA5^{-/-}$ mice), indicating that $MDA5^{-/-}$ mice developed severe heart failure due to virus-induced cardiomyopathy. Thus, MDA5-mediated recognition of EMCV is a prerequisite for triggering antiviral responses as well as for prevention of myocardial dysfunction.

Together, our results demonstrate that RIG-I and MDA5 have essential roles in the recognition of different groups of RNA viruses, as well as in the subsequent production of type-I interferons and pro-inflammatory cytokines. We have found that poly(I:C) and *in vitro* transcribed dsRNA are recognized by MDA5 and RIG-I, respectively; this is in contrast to results from previous *in vitro* studies. RIG-I probably recognizes dsRNA generated over the course of RNA virus replication, as all *in vitro* transcribed dsRNAs tested except for poly(I:C) induced type-I interferons through RIG-I. In contrast, the endogenous ligand of MDA5 remains enigmatic. Moreover, how RIG-I and MDA5 differentially recognize natural dsRNAs is undetermined. Given that the helicase domains of RIG-I and MDA5 bind to dsRNA, analyses of the crystal structures of these domains should to help achieve a better understanding of the molecular mechanisms underlying this differential recognition.

Furthermore, it is still possible that unknown dsRNA-binding proteins also function as direct receptors for viral RNAs.

Finally, the picornavirus family contains several viruses that are pathogenic for humans, including poliovirus, rhinovirus and the virus causing foot-and-mouth-disease. Our studies suggest that human MDA5 and RIG-I also recognize RNA viruses. Thus, identification of therapeutic agents that modify RIG-I or MDA5 may lead to antiviral strategies against selected viruses.

METHODS

Mice, cells and reagents. The generation of $MDA5^{-/-}$ mice is described in the Supplementary Information. $Myd88^{-/-}$, $Tlr3^{-/-}$ and $Trif^{-/-}$ mice have been described previously¹². $Ifnar1^{-/-}$ mice have also been described previously²⁵. $RIG-I^{+/+}$ mice in a 129Sv $\times$ C57BL/6 background were crossed with ICR mice, and the resulting $RIG-I^{+/+}$ mice were further intercrossed. Interbreeding of these $RIG-I^{+/+}$ mice produced healthy and fertile $RIG-I^{-/-}$ offspring, although their number was less than half that of $RIG-I^{+/+}$ progeny (Supplementary Table 1). $RIG-I^{-/-}$ and $RIG-I^{+/+}$ littermate mice were used for *in vivo* experiments. $RIG-I^{-/-}$; $MDA5^{-/-}$ mice in a 129Sv $\times$ C57BL/6 background were lethal at embryonic day 12.5. Additional details regarding cells, reagents and the preparation of *in vitro* transcribed dsRNA are provided in the Supplementary Information.

Viruses. NDV (ref. 3), VSV lacking a variant of M protein (NCP) (ref. 8), influenza virus lacking the NS1 protein (Δ NS1) (ref. 26), JEV (ref. 27) and EMCV (ref. 3) have been described previously. SeV and SeV lacking the V protein (V^{-}) or with mutated C proteins (Cm) were provided by A. Kato²⁸.

Luciferase assay. Wild-type or $RIG-I^{-/-}$; $MDA5^{-/-}$ MEFs were transiently transfected with a reporter construct containing the *Ifnb* promoter together with an empty vector (mock), or *RIG-I* or *MDA5* expression vectors. As an internal control, a *Renilla* luciferase construct was transfected. Transfected cells were untreated or infected with EMCV or SeV Cm (m.o.i. 20) for 24 h. The cells were lysed and subjected to a luciferase assay using a dual-luciferase reporter assay system (Promega) according to the manufacturer's instructions.

Analysis of mice after EMCV infection. Methods for plaque assays, histological analysis and echocardiography are described in the Supplementary Information.

Measurement of cytokine production. Cell culture supernatants were collected and analysed for IFN- β , IFN- α , IL-6 or IL-12p40 production using enzyme-linked immunosorbent assays (ELISAs). ELISA kits for mouse IFN- α and IFN- β were purchased from PBL Biomedical Laboratories, and those for IL-6, IL-12p40 and RANTES were obtained from R&D Systems.

Statistical analysis. Kaplan–Meier plots were constructed and a generalized Wilcoxon test was used to test for differences in survival between control and mutant mice after viral infection. Statistical significance of any differences in cytokine concentration and EMCV titres was determined using Student's t -tests.

Received 30 January; accepted 20 March 2006.

Published online 9 April 2006.

1. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
2. Katze, M. G., He, Y. & Gale, M. Jr. Viruses and interferon: A fight for supremacy. *Nature Rev. Immunol.* **2**, 675–687 (2002).
3. Yoneyama, M. et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
4. Kang, D. C. et al. *mda-5*: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc. Natl Acad. Sci. USA* **99**, 637–642 (2002).
5. Andrejeva, J. et al. The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, *mda-5*, and inhibit its activation of the IFN- β promoter. *Proc. Natl Acad. Sci. USA* **101**, 17264–17269 (2004).
6. Yoneyama, M. et al. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J. Immunol.* **175**, 2851–2858 (2005).
7. Rothenfusser, S. et al. The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *J. Immunol.* **175**, 5260–5268 (2005).
8. Kato, H. et al. Cell type-specific involvement of RIG-I in antiviral response. *Immunity* **23**, 19–28 (2005).
9. Iwasaki, A. & Medzhitov, R. Toll-like receptor control of the adaptive immune responses. *Nature Immunol.* **5**, 987–995 (2004).
10. Beutler, B. Inferences, questions and possibilities in Toll-like receptor signalling. *Nature* **430**, 257–263 (2004).
11. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).

12. Yamamoto, M. *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science* **301**, 640–643 (2003).
13. Kovacsics, M. *et al.* Overexpression of Helicard, a CARD-containing helicase cleaved during apoptosis, accelerates DNA degradation. *Curr. Biol.* **12**, 838–843 (2002).
14. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
15. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κ B and IRF 3. *Cell* **122**, 669–682 (2005).
16. Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN- β signaling. *Mol. Cell* **19**, 727–740 (2005).
17. Meylan, E. *et al.* Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
18. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signaling pathway. *Nature Immunol.* **4**, 491–496 (2003).
19. Sharma, S. *et al.* Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148–1151 (2003).
20. Hemmi, H. *et al.* The roles of two I κ B kinase-related kinases in lipopolysaccharide and double stranded RNA signaling and viral infection. *J. Exp. Med.* **199**, 1641–1650 (2004).
21. Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- α/β gene induction. *Immunity* **13**, 539–548 (2000).
22. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).
23. Chang, T. H., Liao, C. L. & Lin, Y. L. Flavivirus induces interferon-beta gene expression through a pathway involving RIG-I-dependent IRF-3 and PI3K-dependent NF- κ B activation. *Microbes Infect.* **8**, 157–171 (2006).
24. Melchjorsen, J. *et al.* Activation of innate defense against a paramyxovirus is mediated by RIG-I and TLR7 and TLR8 in a cell-type-specific manner. *J. Virol.* **79**, 12944–12951 (2005).
25. Hoshino, K., Kaisho, T., Iwabe, T., Takeuchi, O. & Akira, S. Differential involvement of IFN- β in Toll-like receptor-stimulated dendritic cell activation. *Int. Immunol.* **14**, 1225–1231 (2002).
26. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).
27. Mori, Y. *et al.* Nuclear localization of Japanese encephalitis virus core protein enhances viral replication. *J. Virol.* **79**, 3448–3458 (2005).
28. Kato, A. *et al.* Characterization of the amino acid residues of sendai virus C protein that are critically involved in its interferon antagonism and RNA synthesis down-regulation. *J. Virol.* **78**, 7443–7454 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all colleagues in our laboratory, K. Takeda, T. Shioda, E. Nakayama and K. Kiyotani for helpful discussions, A. Kato, T. Abe, Y. Mori, B. S. Kim and A. Palmenberg for viruses and plasmids, M. Hashimoto for secretarial assistance, and Y. Fujiwara, M. Shiokawa, N. Kitagaki and A. Shibano for technical assistance. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology in Japan, and from the 21st Century Center of Excellence Program of Japan.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.A. (sakira@biken.osaka-u.ac.jp).

LETTERS

A loss-of-function RNA interference screen for molecular targets in cancer

Vu N. Ngo^{1*}, R. Eric Davis^{1*}, Laurence Lamy¹, Xin Yu¹, Hong Zhao¹, Georg Lenz¹, Lloyd T. Lam¹, Sandeep Dave¹, Liming Yang², John Powell² & Louis M. Staudt¹

The pursuit of novel therapeutic agents in cancer relies on the identification and validation of molecular targets. Hallmarks of cancer include self-sufficiency in growth signals and evasion from apoptosis¹; genes that regulate these processes may be optimal for therapeutic attack. Here we describe a loss-of-function screen for genes required for the proliferation and survival of cancer cells using an RNA interference library. We used a doxycycline-inducible retroviral vector for the expression of small hairpin RNAs (shRNAs) to construct a library targeting 2,500 human genes. We used retroviral pools from this library to infect cell lines representing two distinct molecular subgroups of diffuse large B-cell lymphoma (DLBCL), termed activated B-cell-like DLBCL and germinal centre B-cell-like DLBCL. Each vector was engineered to contain a unique 60-base-pair 'bar code', allowing the abundance of an individual shRNA vector within a population of transduced cells to be measured using microarrays of the bar-code sequences. We observed that a subset of shRNA vectors was depleted from the transduced cells after three weeks in culture only if shRNA expression was induced. In activated B-cell-like DLBCL cells, but not germinal centre B-cell-like DLBCL cells, shRNAs targeting the NF- κ B pathway were depleted, in keeping with the essential role of this pathway in the survival of activated B-cell-like DLBCL. This screen uncovered CARD11 as a key upstream signalling component responsible for the constitutive I κ B kinase activity in activated B-cell-like DLBCL. The methodology that we describe can be used to establish a functional taxonomy of cancer and help reveal new classes of therapeutic targets distinct from known oncogenes.

We wished to use RNA interference to perform genetic screens for key regulators of cancer cell proliferation and survival. Because interfering RNAs that target such genes would be toxic, we constructed a retroviral vector for the inducible expression of shRNAs (Fig. 1a). In cells engineered to express the bacterial tetracycline repressor, this vector does not express shRNA until doxycycline is added, and can inducibly knock down the expression of endogenous target genes by 50–70% (Supplementary Fig. S1a). In this vector, we created a library of shRNAs targeting 2,500 human genes, with 3–6 shRNAs per gene. Each shRNA construct was tagged with a different 60-base-pair (bp) bar code to allow us to monitor the abundance of each shRNA vector in a cell population, as previously described^{2,3}.

The experimental strategy of the RNA interference genetic screen is shown in Fig. 1b. The retroviral library is introduced into a cancer cell line and puromycin is used to select stable integrants. The cell population is then divided in two, with one half receiving doxycycline to induce shRNA expression and the other half used as a control cell population. Any shRNA that knocks down the expression of a gene

that is critical for proliferation or survival of the cancer cells will be selectively eliminated from the doxycycline-induced culture. At various time points, genomic DNA is harvested from the two populations and polymerase chain reaction (PCR) is used to amplify the bar-code sequences present in the genomic DNA. Amplified DNAs from the doxycycline-induced and control cultures are fluorescently labelled with different dyes and co-hybridized to a DNA microarray consisting of the bar-code oligonucleotides. The microarray is scanned to reveal the relative abundance of each bar code in the two populations and hence the relative depletion or enrichment of cells expressing a given shRNA.

To test the inducibility of our shRNA vector and the performance of the bar-code screen, we conducted a pilot experiment using the activated B-cell-like DLBCL cell line OCI-Ly3. A pool of 451 shRNA vectors targeting 158 genes was used to infect the lymphoma cells, with each retroviral infection repeated four times to enable statistical analysis. Bar-code representation was monitored in induced versus uninduced cultures after 20 days of shRNA induction and also, as a control, after only 2 days of induction. Figure 2a shows the subset of induced shRNA vectors that were depleted by at least twofold compared with uninduced cultures at day 20 ($P < 0.01$), including shRNAs that target genes regulating the cell cycle (*CDK6*, *BUB1B*), splicing (*PRPF4B*) and NF- κ B signalling (*IKBKB*) (see below)—induced shRNA vectors did not show twofold depletion at day 2. These results demonstrate that the bar-code system can identify shRNA vectors that cause an inducible and time-dependent loss of cancer cells from a culture.

Our primary aim was to identify shRNAs that exhibit 'synthetic lethality' in that they inhibit the proliferation and survival of only certain cancer subgroups due to molecular differences between the subgroups. To this end, our genetic screen interrogated growth and survival pathways in two molecular subgroups of DLBCLs—activated B-cell-like DLBCLs and germinal centre B-cell-like DLBCLs—that differ profoundly in gene expression, genomic abnormalities and clinical outcome^{4,5}. Most importantly, activated B-cell-like DLBCLs, but not germinal centre B-cell-like DLBCLs, rely on constitutive NF- κ B signalling for survival^{6,7}. Genetic screens using a subset of the shRNA library (1,854 shRNA vectors targeting 683 genes) were performed in two activated B-cell-like DLBCL cell lines (OCI-Ly3 and OCI-Ly10) and in two germinal centre B-cell-like DLBCL cell lines (OCI-Ly7 and OCI-Ly19). Altogether, 17 shRNA vectors targeting 15 genes were significantly depleted ($P < 0.01$) from induced cultures of at least two cell lines (Supplementary Fig. S2).

Notably, all of the shRNAs that were selectively toxic for activated B-cell-like DLBCLs targeted genes that regulate the NF- κ B pathway, including a shRNA targeting the central kinase in this pathway,

¹Metabolism Branch, Center for Cancer Research, National Cancer Institute, and ²Bioinformatics and Molecular Analysis Section, Computational Bioscience and Engineering Laboratory, CIT, National Institutes of Health, Bethesda, Maryland 20892, USA.

*These authors contributed equally to this work.

IKBKB (Fig. 2b; see also Supplementary Fig. S2). Also included were two separate shRNAs for *CARD11*, which is required for NF- κ B activation by antigen receptor signalling in B and T lymphocytes, as well as individual shRNAs targeting *MALT1* and *BCL10*, two proteins that cooperate with *CARD11* in NF- κ B signalling⁸. The ability of these shRNAs to reduce expression of their respective messenger RNAs was confirmed by quantitative reverse transcription-PCR (Q-PCR) analysis (Supplementary Fig. S1b, c). To expand our repertoire of shRNAs for functional validation studies, we identified two additional *CARD11* shRNAs and two additional *MALT1* shRNAs that effectively decreased expression of their cognate mRNA and protein (Supplementary Fig. S1b, d).

To confirm and extend the results from the bar-code screen, we established a second toxicity assay in which we could follow the fate of cells acutely transduced with a single shRNA over time. To do this, we cloned *CARD11* shRNA 1 into a modified vector that co-expresses green fluorescent protein (GFP). After retroviral transduction, we monitored the proportion of GFP⁺ cells versus GFP⁻ cells over time as a measure of the toxicity of the co-expressed shRNA. Experiments were conducted in four activated B-cell-like DLBCL cell lines, four germinal centre B-cell-like DLBCL cell lines and three cell lines representing a third type of DLBCL termed primary mediastinal B-cell lymphoma (PMBL)^{9,10} (Supplementary Fig. S3). In all four activated B-cell-like DLBCL cell lines, the proportion of GFP⁺ cells decreased over time, indicating a toxic effect of the *CARD11* shRNA, whereas no toxicity was observed in the germinal centre B-cell-like DLBCL and PMBL cell lines (Fig. 3a, b). Additional *CARD11* and *MALT1* shRNAs also demonstrated toxicity for OCI-Ly3 cells in this assay (Fig. 3c). Notably, the *CARD11* shRNA showed no toxicity for

PMBL cell lines despite the fact that this lymphoma type requires constitutive activity of the NF- κ B pathway for survival⁷. The selective toxicity of *CARD11* shRNAs for activated B-cell-like DLBCLs suggests that the molecular mechanisms causing constitutive activity of the NF- κ B pathway differ between activated B-cell-like DLBCL and PMBL.

To test whether these shRNAs inhibited the NF- κ B pathway in activated B-cell-like DLBCLs, we used a cell-based reporter assay for I κ B kinase (IKK) activity based on a fusion protein between I κ B α and *Photinus* luciferase⁷. When IKK is inhibited, this reporter is not phosphorylated and is consequently not degraded in the proteasome, resulting in a rise in luciferase activity⁷. In activated B-cell-like DLBCL cells, retroviral expression of various shRNAs for *IKBKB*, *CARD11*, *MALT1* and *BCL10* produced a rise in the I κ B α luciferase reporter, but shRNAs targeting these genes did not affect this reporter in germinal centre B-cell-like DLBCL cells (Fig. 4a). Control shRNAs targeting *POU2F2*, *YY1*, *WNK1* or *GFP* did not affect this reporter and, as expected, a shRNA vector targeting *Photinus* luciferase decreased the reporter signal (Fig. 4a).

We next used DNA microarrays to investigate whether the target genes of the NF- κ B pathway were downregulated in response to the *CARD11* shRNA. OCI-Ly3 cells were transduced with the inducible *CARD11* shRNA vector, and gene expression was compared between doxycycline-induced cultures and parallel, uninduced cultures. A previously defined set of NF- κ B target genes in activated B-cell-like DLBCLs⁷ displayed down-modulated expression upon *CARD11* shRNA induction (Fig. 4b). The secretion of IL-6, one of the targets of NF- κ B, was also decreased when *CARD11* shRNA was expressed (Fig. 4c). Thus, inhibition of *CARD11* in activated B-cell-like

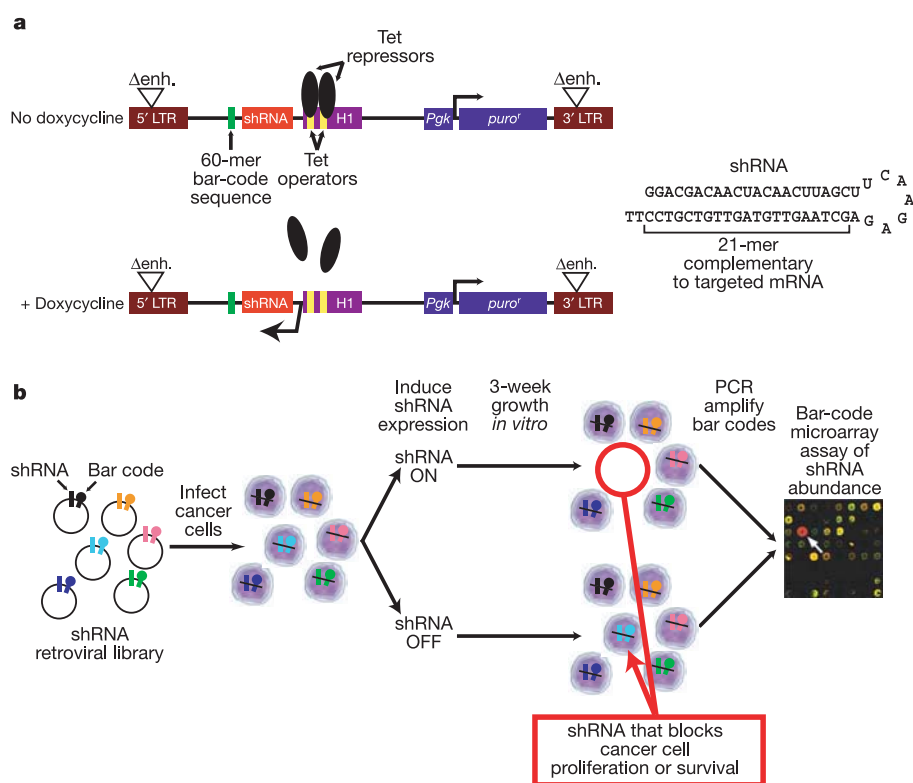


Figure 1 | Inducible shRNA library screen for genes controlling cancer cell proliferation and survival. **a**, Structure of the inducible shRNA-expressing retroviral vector after integration into genomic DNA of infected cells. Two tetracycline repressor (TETR) binding sites (Tet operators) were inserted into the histone H1 promoter, which drives shRNA expression. In cell lines expressing the TETR, shRNA from this vector is only expressed upon addition of doxycycline, which dissociates the TETR from the H1 promoter. A random 60-bp molecular bar-code oligonucleotide was cloned adjacent to

each shRNA, and the association of a particular bar-code sequence with a particular shRNA was determined by sequencing each vector in the library. Also shown is a diagram of the shRNA structure, consisting of a 21-base sequence complementary to a targeted mRNA and a 9-base loop. LTR, long terminal repeat; *puro'*, puromycin resistance gene; *Pgk*, phosphoglycerate kinase promoter; Δ enh., deletion of LTR promoter sequences. **b**, Achilles' heel loss-of-function genetic screen for genes required for cancer cell proliferation or survival. (See text for details.)

DLBCL cells inhibits IKK and decreases expression of NF- κ B target genes.

The present study has uncovered a previously unsuspected role for CARD11 signalling in the pathogenesis of DLBCL. CARD11 signalling in activated B-cell-like DLBCL seems to engage MALT1 and BCL10 to activate IKK, in keeping with previous studies⁸. However, because *CARD11* shRNAs were consistently more toxic than *MALT1* shRNAs, it is possible that CARD11 also signals to other pathways in activated B-cell-like DLBCL that do not require MALT1 (ref. 11). At present, the molecular mechanisms accounting for the constitutive CARD11 signalling in activated B-cell-like DLBCL are unknown. One possibility is that activated B-cell-like DLBCLs harbour oncogenic lesions that alter the expression and/or function of CARD11, BCL10 or MALT1. In gastric mucosa associated lymphoid tissue lymphoma, *MALT1* and *BCL10* have been implicated in recurrent chromosomal translocations¹². The most frequent of these translocations—t(11;18), involving *MALT1*—is not present in activated B-cell-like DLBCL cell lines or primary tumours (data not shown). Moreover, activated B-cell-like DLBCLs do not have higher levels of *MALT1* or *BCL10* expression than germinal centre

B-cell-like DLBCLs, arguing against the existence of translocations that upregulate the expression of these genes in activated B-cell-like DLBCL.

Alternatively, CARD11-dependent IKK activation in activated B-cell-like DLBCL may be a physiological attribute of the type of B-cell subpopulation from which activated B-cell-like DLBCL is derived. Mice with mutations or deletions in *CARD11*, *MALT1* or *BCL10* are severely deficient in two B-cell subpopulations—the marginal zone B cell and the mature B1 B cell—but have little or no deficiency in mature follicular B2 cells^{11,13–20}. Although the cell of origin of activated B-cell-like DLBCLs has not been clearly defined, it could conceivably derive from a B-cell subset that depends critically on CARD11/BCL10/MALT1 for survival and/or self renewal. Furthermore, primary tumour biopsies of activated B-cell-like DLBCLs show consistently higher expression of *CARD11* mRNA than germinal centre B-cell-like DLBCL biopsies (~2-fold; $P < 10^{-11}$) and PMBL biopsies (~5-fold; $P < 10^{-18}$) (Supplementary Fig. S4), which might contribute to the CARD11 dependence of activated B-cell-like DLBCL, as experimental overexpression of CARD11 can activate IKK in a BCL10-dependent manner^{21–23}.

CARD11 and associated molecules are attractive therapeutic targets for activated B-cell-like DLBCL. Inhibition of CARD11 signalling would, in theory, have consequences only in the lymphoid system, as CARD11 expression is restricted to the lymphoid system

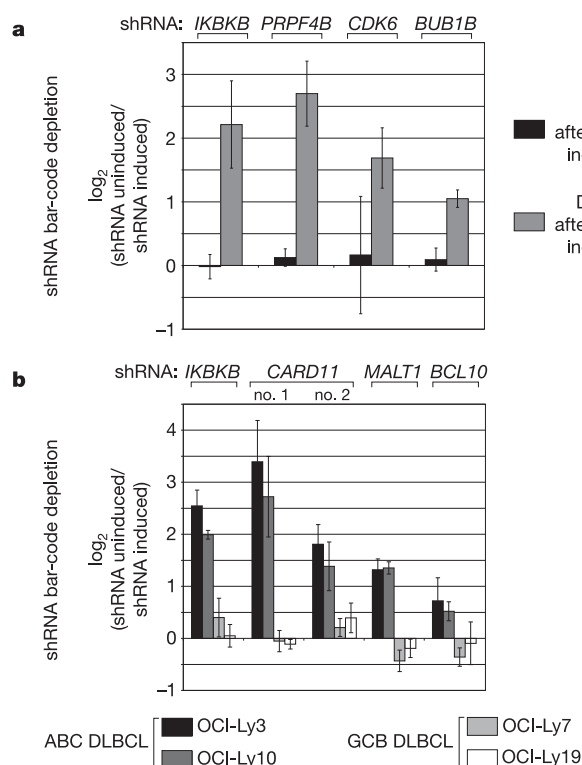


Figure 2 | Identification of shRNAs that block the proliferation or survival of lymphoma cell lines. Shown are the relative fluorescent signals from bar-code microarray experiments comparing shRNA induced versus uninduced cells. A higher value along the y axis indicates selective depletion of a shRNA from the induced cell population. **a**, A pilot genetic screen with 459 shRNA vectors in the activated B-cell-like DLBCL cell line OCI-Ly3. Shown are shRNA vectors that were depleted >2-fold in infected cells at day 20, but not at day 2, after doxycycline induction (P values <0.01; paired t -test). **b**, Genetic screens using 1,854 shRNA vectors targeting 683 genes performed in two activated B-cell-like (ABC) DLBCL cell lines (OCI-Ly3 and OCI-Ly10) and two germinal centre B-cell-like (GCB) DLBCL cell lines (OCI-Ly7 and OCI-Ly19). Relative bar-code abundance was measured at day 21 after shRNA induction. The shRNA vectors targeting *IKKB*, *CARD11* and *MALT1* were depleted >2-fold in both induced activated B-cell-like DLBCL cell populations ($P < 0.01$; paired t -test), but not in the induced germinal centre B-cell-like DLBCL cells. A *BCL10* shRNA vector was also preferentially depleted in induced activated B-cell-like DLBCL cell populations ($P < 0.05$). Data are the mean $\pm$ s.d. of four independent retroviral transductions of the shRNA library.

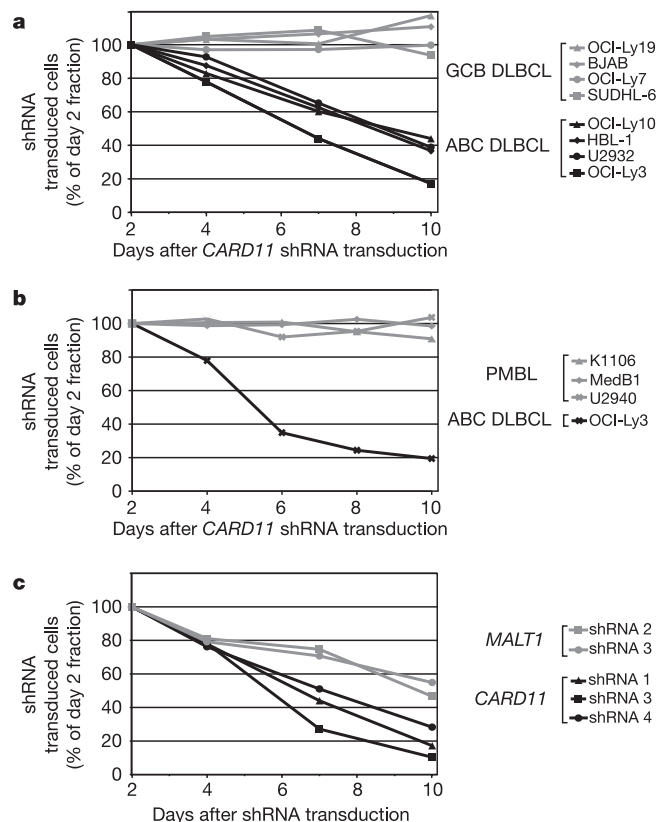


Figure 3 | Toxicity of *CARD11* and *MALT1* shRNAs for activated B-cell-like DLBCL cell lines. Lymphoma cell lines were transduced with a retroviral vector that co-expresses a shRNA and GFP, and the fraction of GFP⁺ cells was measured at the indicated times by FACS. The toxicity of a shRNA was evident by a reduction in the GFP⁺ fraction over time as compared to the GFP⁺ fraction at day 2 after retroviral transduction (see Methods for details). Data are representative of at least two independent experiments. **a**, Toxicity of *CARD11* shRNA 1 for activated B-cell-like DLBCL but not germinal centre B-cell-like DLBCL cell lines. **b**, Toxicity of *CARD11* shRNA 1 for the activated B-cell-like DLBCL cell line OCI-Ly3 but not PMBL cell lines. **c**, Toxicity of multiple *CARD11* and *MALT1* shRNAs for the activated B-cell-like DLBCL cell line OCI-Ly3.

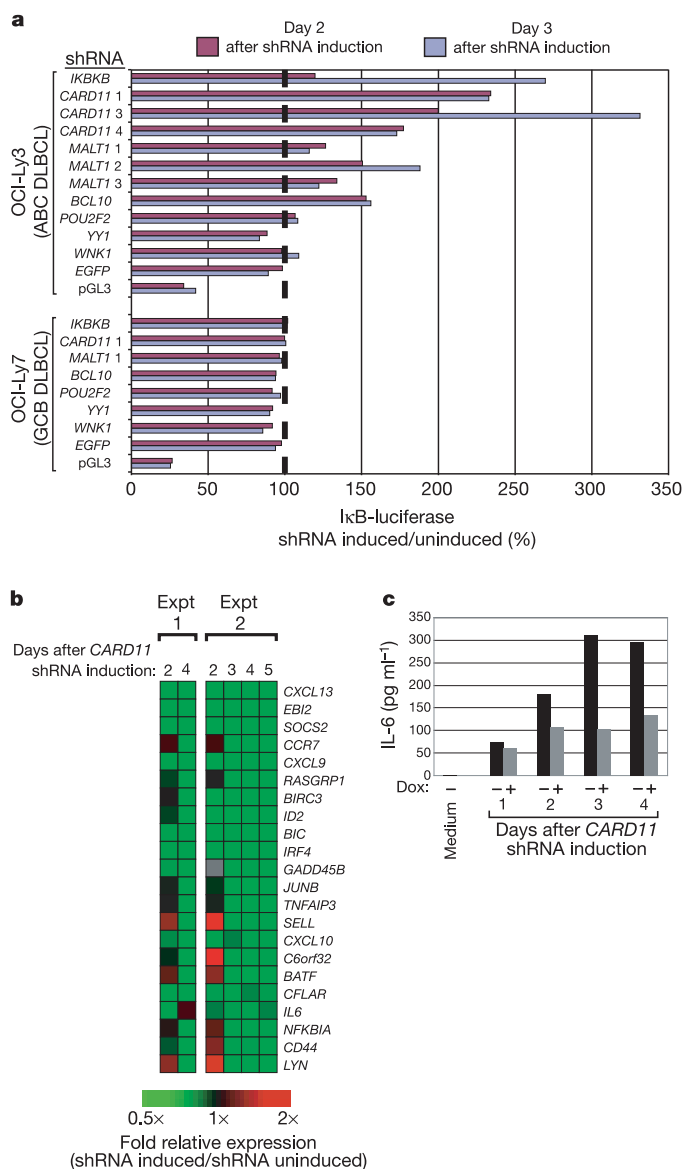


Figure 4 | Role of CARD11, MALT1 and BCL10 in NF- κ B signalling in activated B-cell-like DLBCL. **a**, IKK activity in activated B-cell-like DLBCL depends on CARD11, MALT1 and BCL10. IKK reporter cell lines bearing an IKB α luciferase reporter were transduced with retroviral vectors for the indicated shRNAs. Luciferase activity was measured at the indicated time points and the activity in shRNA induced cells is shown as a percentage of the activity in uninduced cells (see Methods). Data are representative of at least two independent experiments. **b**, CARD11 signalling activates NF- κ B target gene expression in activated B-cell-like DLBCL cells. Expression of CARD11 shRNA 1 was induced in OCI-Ly3 cells for the indicated times and total RNA was harvested from induced cultures and from parallel uninduced cultures. DNA microarrays were used to determine relative gene expression in induced and uninduced cells. Downregulation of NF- κ B target genes in two independent experiments is indicated in green according to the colour scale shown. **c**, IL-6 secretion is downstream of CARD11 signalling in activated B-cell-like DLBCL. CARD11 shRNA 1 was induced in OCI-Ly3 cells for the indicated times. The concentration of IL-6 was measured in the supernatant of induced cells and in parallel uninduced cells.

and CARD11 mutant mice only have lymphoid abnormalities⁸. In normal lymphocytes, MALT1 can activate IKK by ubiquitinating the IKK- γ subunit; it can also induce TRAF6 to function as an IKK- γ ubiquitin ligase^{24,25}. New therapeutics targeting these enzymatic reactions could prove useful in the therapy of DLBCLs that depend on the CARD11 pathway for survival.

Previous RNA interference screens in cancer have been 'positive'

screens that rely on the ability of an interfering RNA to rescue a cell from some cytotoxic or cytostatic influence^{3,26,27}. Such screens are ideal for uncovering tumour suppressor genes, but are not readily adapted for the identification of oncogenes that promote the malignant phenotype. By contrast, the inducible nature of our shRNA vector allowed us to perform a 'negative' screen in which interfering RNAs produce a cytotoxic or cytostatic phenotype, thereby uncovering oncogenic pathways that promote malignancy.

The 'Achilles' heel' methodology outlined in the present report can be extended to create a functional taxonomy of cancer that should be particularly relevant for the development of new therapeutic agents. In this scenario, cancers would be classified on the basis of which regulatory proteins/pathways promote proliferation or prevent cell death. This functional classification of cancer would be expected, in some cases, to cut across diagnostic categories provided by pathological examination or gene expression profiling. The functionally critical pathways identified in such genetic screens could also guide a focused resequencing effort in cancer genomes to search for mutated genes that lie within these pathways.

An equally exciting possibility is that Achilles' heel screens will uncover regulatory pathways that are not directly activated by underlying oncogenic events but that are nevertheless required for cancer cell proliferation or survival. Cancers may arise from a stage of normal cellular differentiation in which a particular signalling pathway is used to prevent cell death or promote proliferation, and the derived cancer cell may retain a dependence upon that pathway. Given the diversity of physiological regulators of apoptosis and proliferation, many of which are cell-type specific, future Achilles' heel screens are likely to extend the realm of molecular targets in cancer.

METHODS

See Supplementary Methods for detailed experimental methods.

Construction of an inducible shRNA retroviral expression library. The inducible shRNA expression plasmid pRSMX was derived from pRetroSuper²⁸ by modifying the H1 promoter with binding sites for the bacterial tetracycline repressor, as described²⁹. A library of random 60-mer bar-code sequences was generated in pRSMX by cloning into *MfeI/XhoI* sites immediately 3' of the shRNA cloning site. Cloned shRNA and bar-code sequences were verified by sequencing. shRNA oligonucleotides (Invitrogen) were chosen to have 21-bp complementarity with the target gene, following previously described design rules³⁰, and had <15-bp identity to any other RefSeq annotated mRNA sequence.

Preparation of doxycycline-inducible cell lines. Each cell line used in the bar-code screen was first transduced with a feline endogenous virus expressing the ecotropic retroviral receptor and then secondarily infected with an ecotropic retrovirus expressing the bacterial tetracycline repressor (TETR).

Bar-code shRNA screen. All of the shRNA expression constructs from a single 96-well plate were pooled into a single plasmid prep and, typically, five such plasmid pools were combined and used to generate a high-titre ecotropic retroviral stock. After infection of TETR-expressing lymphoma cells and puromycin selection, doxycycline (20 ng ml⁻¹) was added to induce shRNA expression in half of the culture. Bar-code sequences were amplified from genomic DNA and hybridized to bar-code microarrays essentially as described³. The sequences of the effective shRNAs are provided in Supplementary Table 1.

Survival assay. A variant of pRSMX was created in which the puromycin resistance (*puro*^r) gene was replaced with a fusion gene between *puro*^r and GFP. shRNA-expressing retroviruses prepared from this vector were used to infect lymphoma cell lines (lacking the TETR), and the fraction of cells that were GFP⁺ was measured over time by FACS. For each time point and shRNA, the observed GFP⁺ fractions were normalized to the GFP⁺ fraction in parallel cultures transduced with a control shRNA, and compared to the initial GFP⁺ fraction 2 days after retroviral transduction.

Received 5 December 2005; accepted 2 March 2006.

Published online 29 March 2006.

1. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* 100, 57–70 (2000).
2. Paddison, P. J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* 428, 427–431 (2004).

3. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
4. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
5. Rosenwald, A. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346**, 1937–1947 (2002).
6. Davis, R. E., Brown, K. D., Siebenlist, U. & Staudt, L. M. Constitutive nuclear factor κ B activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J. Exp. Med.* **194**, 1861–1874 (2001).
7. Lam, L. T. *et al.* Small molecule inhibitors of I κ B-kinase are selectively toxic for subgroups of diffuse large B cell lymphoma defined by gene expression profiling. *Clin. Cancer Res.* **11**, 28–40 (2005).
8. Thome, M. CARMA1, BCL-10 and MALT1 in lymphocyte development and activation. *Nature Rev. Immunol.* **4**, 348–359 (2004).
9. Rosenwald, A. *et al.* Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J. Exp. Med.* **198**, 851–862 (2003).
10. Savage, K. J. *et al.* The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood* **102**, 3871–3879 (2003).
11. Ruland, J., Duncan, G. S., Wakeham, A. & Mak, T. W. Differential requirement for Malt1 in T and B cell antigen receptor signaling. *Immunity* **19**, 749–758 (2003).
12. Isaacson, P. G. & Du, M. Q. MALT lymphoma: from morphology to molecules. *Nature Rev. Cancer* **4**, 644–653 (2004).
13. Zhang, Q. *et al.* Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). *Nature Genet.* **22**, 63–68 (1999).
14. Ruefli-Brasse, A. A., French, D. M. & Dixit, V. M. Regulation of NF- κ B-dependent lymphocyte activation and development by paracaspase. *Science* **302**, 1581–1584 (2003).
15. Ruland, J. *et al.* Bcl10 is a positive regulator of antigen receptor-induced activation of NF- κ B and neural tube closure. *Cell* **104**, 33–42 (2001).
16. Xue, L. *et al.* Defective development and function of Bcl10-deficient follicular, marginal zone and B1 B cells. *Nature Immunol.* **4**, 857–865 (2003).
17. Newton, K. & Dixit, V. M. Mice lacking the CARD of CARMA1 exhibit defective B lymphocyte development and impaired proliferation of their B and T lymphocytes. *Curr. Biol.* **13**, 1247–1251 (2003).
18. Egawa, T. *et al.* Requirement for CARMA1 in antigen receptor-induced NF- κ B activation and lymphocyte proliferation. *Curr. Biol.* **13**, 1252–1258 (2003).
19. Hara, H. *et al.* The MAGUK family protein CARD11 is essential for lymphocyte activation. *Immunity* **18**, 763–775 (2003).
20. Jun, J. E. *et al.* Identifying the MAGUK protein Carma-1 as a central regulator of humoral immune responses and atopy by genome-wide mouse mutagenesis. *Immunity* **18**, 751–762 (2003).
21. McAllister-Lucas, L. M. *et al.* Bim1, a MAGUK family member linking protein kinase C activation to Bcl10-mediated NF- κ B induction. *J. Biol. Chem.* **276**, 30589–30597 (2001).
22. Gaide, O. *et al.* Carma1, a CARD-containing binding partner of Bcl10, induces Bcl10 phosphorylation and NF- κ B activation. *FEBS Lett.* **496**, 121–127 (2001).
23. Bertin, J. *et al.* CARD11 and CARD14 are novel caspase recruitment domain (CARD)/membrane-associated guanylate kinase (MAGUK) family members that interact with BCL10 and activate NF- κ B. *J. Biol. Chem.* **276**, 11877–11882 (2001).
24. Zhou, H. *et al.* Bcl10 activates the NF- κ B pathway through ubiquitination of NEMO. *Nature* **427**, 167–171 (2004).
25. Sun, L., Deng, L., Ea, C. K., Xia, Z. P. & Chen, Z. J. The TRAF6 ubiquitin ligase and TAK1 kinase mediate IKK activation by BCL10 and MALT1 in T lymphocytes. *Mol. Cell* **14**, 289–301 (2004).
26. Kolfschoten, I. G. *et al.* A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* **121**, 849–858 (2005).
27. Westbrook, T. F. *et al.* A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837–848 (2005).
28. Brummelkamp, T. R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243–247 (2002).
29. van de Wetering, M. *et al.* Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. *EMBO Rep.* **4**, 609–615 (2003).
30. Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nature Biotechnol.* **22**, 326–330 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. V.N.N. was also supported by a Damon Runyon-Walter Winchell Cancer Research Foundation Fellowship.

Author Information The microarray data discussed in this publication have been deposited in the Gene Expression Omnibus of NCBI (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO series accession number GSE3896. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.M.S. (lstaedt@mail.nih.gov).

RNAi-mediated gene silencing in non-human primates

Tracy S. Zimmermann¹, Amy C. H. Lee², Akin Akinc¹, Birgit Bramlage³, David Bumcrot¹, Matthew N. Fedoruk², Jens Harborth¹, James A. Heyes², Lloyd B. Jeffs², Matthias John³, Adam D. Judge², Kieu Lam², Kevin McClintock², Lubomir V. Nechev¹, Lorne R. Palmer², Timothy Racie¹, Ingo Röhl³, Stephan Seiffert³, Sumi Shanmugam¹, Vandana Sood², Jürgen Soutschek³, Ivanka Toudjarska¹, Amanda J. Wheat², Ed Yaworski², William Zedalis¹, Victor Koteliansky¹, Muthiah Manoharan¹, Hans-Peter Vornlocher³ & Ian MacLachlan²

The opportunity to harness the RNA interference (RNAi) pathway to silence disease-causing genes holds great promise for the development of therapeutics directed against targets that are otherwise not addressable with current medicines^{1,2}. Although there are numerous examples of *in vivo* silencing of target genes after local delivery of small interfering RNAs (siRNAs)^{3–5}, there remain only a few reports of RNAi-mediated silencing in response to systemic delivery of siRNA^{6–8}, and there are no reports of systemic efficacy in non-rodent species. Here we show that siRNAs, when delivered systemically in a liposomal formulation, can silence the disease target apolipoprotein B (ApoB) in non-human primates. *APOB*-specific siRNAs were encapsulated in stable nucleic acid lipid particles (SNALP) and administered by intravenous injection to cynomolgus monkeys at doses of 1 or 2.5 mg kg⁻¹. A single siRNA injection resulted in dose-dependent silencing of *APOB* messenger RNA expression in the liver 48 h after administration, with maximal silencing of >90%. This silencing effect occurred as a result of *APOB* mRNA cleavage at precisely the site predicted for the RNAi mechanism. Significant reductions in ApoB protein, serum cholesterol and low-density lipoprotein levels were observed as early as 24 h after treatment and lasted for 11 days at the highest siRNA dose, thus demonstrating an immediate, potent and lasting biological effect of siRNA treatment. Our findings show clinically relevant RNAi-mediated

gene silencing in non-human primates, supporting RNAi therapeutics as a potential new class of drugs.

ApoB is expressed predominantly in the liver and jejunum, and is an essential protein for the assembly and secretion of very-low-density lipoprotein (VLDL) and low-density lipoprotein (LDL), which are required for the transport and metabolism of cholesterol⁹. As a large, lipid-associated protein, ApoB is not accessible to targeting with conventional therapies, but it is a highly relevant and validated disease target. Elevated ApoB and LDL levels are correlated with increased risk of coronary artery disease, and inadequate control of LDL-cholesterol after acute coronary syndromes results in increased risk of recurrent cardiac events or death^{10,11}. Approaches targeting ApoB with second-generation antisense oligonucleotides have progressed to pre-clinical and clinical studies¹². Despite progress in the management of hypercholesterolaemia using HMG-CoA reductase inhibitors and other drugs that affect dietary cholesterol, there remains a significant need for new therapeutic approaches.

We have previously demonstrated silencing of *ApoB* in rodents using cholesterol-conjugated siRNAs⁶. In the current study, we used a liposomal formulation of SNALP to evaluate systemic delivery of siRNA directed towards *APOB*. Preliminary evaluations were conducted in mice. Whereas administration of the *ApoB*-specific siRNA siApoB-1, without formulation or chemical conjugation, at doses higher than 50 mg kg⁻¹ was previously shown to have no

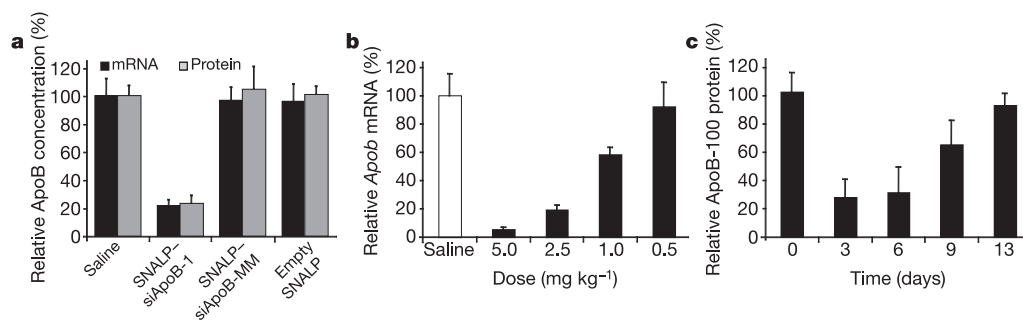


Figure 1 | SNALP-siRNA-mediated silencing of murine *ApoB* is potent, specific, dose-dependent and long-lasting. a, Liver *ApoB* mRNA levels normalized to *Gapdh* mRNA and serum ApoB-100 protein levels measured two days after single i.v. injections of saline, SNALP-siApoB-1 (1 mg kg⁻¹), mismatched SNALP-siApoB-MM (1 mg kg⁻¹) or empty SNALP vesicles (25 mg kg⁻¹) ($n = 5$ per group). **b**, Liver *ApoB* mRNA levels normalized to

Gapdh mRNA, assessed three days after i.v. administration of saline or 5, 2.5, 1 or 0.5 mg kg⁻¹ SNALP-siApoB-2 ($n = 4$ per group). **c**, Serum ApoB-100 levels after i.v. administration of either saline or 2.5 mg kg⁻¹ SNALP-siApoB-2 ($n = 6$ per group). Serum ApoB-100 levels for SNALP-siApoB-2-treated animals are relative to the saline-treated group for the same time point. Data show mean $\pm$ s.d.

¹Alnylam Pharmaceuticals Inc., 300 Third Street, Cambridge, Massachusetts 02142, USA. ²Protiva Biotherapeutics Inc., 100-3480 Gilmore Way, Burnaby, British Columbia V5G 4Y1, Canada. ³Alnylam Europe AG, Fritz-Hornschuch-Str. 9, 95326 Kulmbach, Germany.

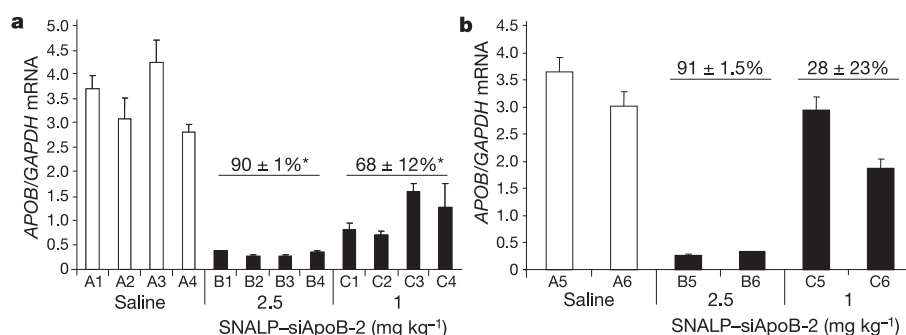


Figure 2 | Systemic silencing of APOB mRNA in non-human primates.

a, b, Liver APOB mRNA levels for 12 biopsies (three isolated from each of four liver lobes) were quantified relative to GAPDH mRNA either 48 h (**a**, $n = 4$ animals per group) or 11 days (**b**, $n = 2$) after treatment with SNALP-siApoB-2. Data shown are mean APOB/GAPDH mRNA

levels $\pm$ s.d. for each animal. Mean values ($\pm$ s.d.) of the per cent APOB mRNA reduction relative to the saline treatment group are shown above each group. Asterisks indicate statistical significance compared with the saline-treated group ($P < 0.005$; ANOVA).

in vivo silencing activity⁶, ~80% silencing of liver *ApoB* mRNA and ApoB-100 protein was achieved with a single 1 mg kg⁻¹ dose of SNALP-formulated siApoB-1 (Fig. 1a). In contrast, no detectable reduction was observed with a SNALP-formulated mismatched siRNA (siApoB-MM) or empty SNALP vesicles, indicating that silencing is specific to the siRNA and is not caused by the liposomal carrier. This silencing effect of SNALP-formulated siRNA represents more than a 100-fold improvement in potency compared with systemic administration of cholesterol-conjugated siApoB-1 (chol-siApoB-1) (Supplementary Fig. 1). Moreover, liposomal formulation of siRNA seems to be a general strategy for silencing hepatocyte targets, as demonstrated in mice for coagulation factor VII, green fluorescent protein and cyclophilin B (A.A., R. Constien and M.N.F., unpublished results).

As siApoB-1 was originally designed to be cross-reactive to both mouse and human ApoB genes, and we planned to conduct RNAi studies in non-human primates, a second ApoB-specific siRNA, siApoB-2, was designed to be cross-reactive with mouse, human and cynomolgus monkey ApoB genes. siApoB-2 was also selected on the basis of *in vitro* gene silencing activity and the absence of immunostimulatory activity (data not shown). Murine studies showed that encapsulated siApoB-2 showed a dose-dependent reduction in *ApoB* mRNA, with >90% silencing achieved at the highest (5 mg kg⁻¹) dose (Fig. 1b). After a single 2.5 mg kg⁻¹ dose of SNALP-siApoB-2, 80% silencing of liver *ApoB* mRNA was associated with a 72% reduction in serum ApoB-100 protein. The silencing effect was detected for up to nine days, and was followed by recovery to normal protein levels by day 13 after treatment (Fig. 1c).

To address the therapeutic potential of this systemic RNAi approach, we evaluated the pharmacokinetics, efficacy and safety of SNALP-formulated siApoB-2 in cynomolgus monkeys. We first determined the circulating half-life of SNALP-siApoB-2 in plasma samples collected from cynomolgus monkeys ($n = 2$) receiving a single 2.5 mg kg⁻¹ intravenous (i.v.) injection of the siRNA. An elimination half-life of 72 min was measured for the siRNA (Supplementary Fig. 2), compared with a 38-min half-life in mice (Supplementary Fig. 3a).

To evaluate efficacy, cynomolgus monkeys were treated with saline or SNALP-formulated siApoB-2 at doses of 1 or 2.5 mg kg⁻¹ ($n = 6$ per group). siApoB-2 treatment was associated with a clear and statistically significant dose-dependent gene-silencing effect on cynomolgus liver APOB mRNA. Forty-eight hours after treatment, APOB mRNA was reduced by 68 $\pm$ 12% (mean $\pm$ s.d., $n = 4$, $P = 0.004$) and 90 $\pm$ 1% ($n = 4$, $P = 0.002$) for the 1 mg kg⁻¹ and 2.5 mg kg⁻¹ groups, respectively (Fig. 2a). Gene silencing was found to be consistent across the liver and correlated with detectable tissue levels of siApoB-2 (Supplementary Fig. 4). We also confirmed this APOB mRNA silencing to be mediated by RNAi, as demonstrated by

5' rapid amplification of cDNA ends (RACE) analysis and identification of the predicted cleavage site, exactly ten nucleotides from the 5' end of the antisense strand of siApoB-2 (Supplementary Fig. 5). Notably, APOB mRNA silencing was maintained for 11 days after the single 2.5 mg kg⁻¹ treatment, with APOB mRNA levels still reduced by 91 $\pm$ 1.5% (Fig. 2b). Monkeys treated with the 1 mg kg⁻¹ dose showed varying degrees of recovery from ApoB silencing at the day 11 time point. Although APOB mRNA was efficiently silenced in the liver, SNALP-siApoB-2 showed no silencing of APOB expressed in the jejunum (Supplementary Fig. 6), consistent with the absence of significant biodistribution of SNALP-formulated siRNAs to intestinal tissues in mice (Supplementary Fig. 3b).

The degree and persistence of RNAi-mediated silencing observed in cynomolgus monkeys far exceeds the results obtained with rodents. The lasting RNAi-mediated effects *in vivo* are consistent with observed long-lasting silencing by siRNAs in other studies^{13,14}, and the longer duration observed in primates may relate to species differences in the efficiency and stability of the RNA-induced

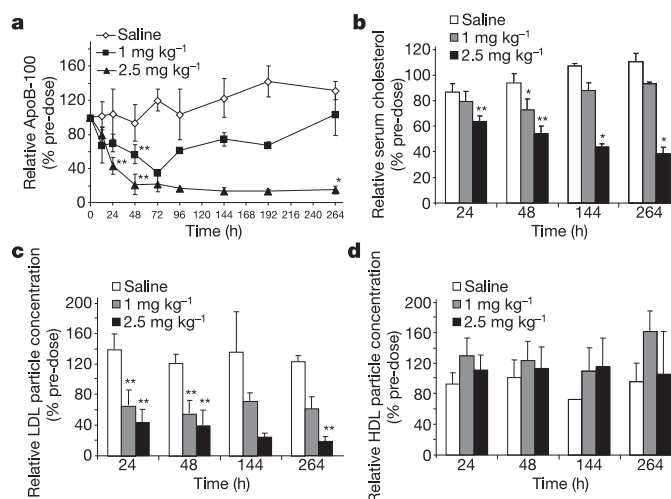


Figure 3 | Phenotypic effects of RNAi-mediated silencing of APOB mRNA in non-human primates. **a–d,** Serial plasma samples were obtained from cynomolgus monkeys treated with saline or 1 or 2.5 mg kg⁻¹ SNALP-siApoB-2, and measured for ApoB-100 (**a**), total serum cholesterol (**b**), LDL (**c**) and HDL (**d**) levels. Data show levels as a percentage of pre-dose values and are expressed as mean $\pm$ s.d. Data sets collected at 0, 12, 24 and 48 h have a group size of six, and data sets collected at later time points have a group size of two. Data points marked with asterisks are statistically significant compared with saline-treated animals (* $P < 0.05$, ** $P < 0.005$; ANOVA).

silencing complex (RISC), the mitotic state of hepatocytes and/or the tissue stability of the siRNA.

The expected biological effects resulting from *APOB* mRNA silencing include reduction in the blood levels of ApoB-100 protein, total cholesterol and LDL. To evaluate the kinetics of these downstream effects, we analysed plasma sampled serially from individual monkeys before and during the 11-day time course of the single-dose siApoB-2 study. Plasma ApoB-100 protein levels were reduced as early as 12 h after administration of 1 or 2.5 mg kg⁻¹ SNALP-siApoB-2, reaching nadirs of 35 ± 2% and 22 ± 9% of pre-treatment levels, respectively, 72 h after treatment (Fig. 3a). Animals that received the higher siRNA dose maintained a marked reduction in ApoB protein between 2 and 11 days after treatment, consistent with the lasting effect on mRNA silencing. Monkeys that received the lower siRNA dose showed an intermediate degree of ApoB protein reduction that returned to pre-dose levels by day 11, consistent with the observed recovery in *APOB* mRNA.

Serum cholesterol levels were similarly reduced, in a dose-dependent manner and with comparable kinetics (Fig. 3b). The maximum cholesterol reduction of 62 ± 5.5% ($n = 2$, $P = 0.006$) observed for the high dose siRNA group would be considered clinically significant for patients with hypercholesterolaemia, and exceeds levels of cholesterol reduction reported clinically for currently approved cholesterol-lowering drugs.

Administration of SNALP-siApoB-2 also resulted in dramatic and rapid dose-dependent reduction in the ApoB-containing lipoprotein particle LDL. Reduction in LDL relative to pre-dose levels was observed as early as 24 h after treatment for both doses of SNALP-siApoB-2 (Fig. 3c). In contrast, there were no significant changes in circulating levels of the non-ApoB-containing high-density lipoprotein particle (HDL, Fig. 3d). The reduction in LDL persisted over the 11-day study for both siApoB-2 treatment groups, with a maximum 82 ± 7% decrease compared to pre-treatment levels observed for the high-dose group at day 11 ($n = 2$, $P = 0.003$). The time required for the biological effects to return to pre-dose levels was not determined for the high-dose group because the endpoint for this study was defined using rodent data, which indicated a faster rate of recovery. The rapid onset and lasting effect on lipoprotein metabolism suggest that siRNAs targeting *APOB* may be a valuable therapeutic strategy for achieving plaque stabilization in acute coronary syndromes^{10,11}, as HMG-CoA reductase inhibitors can require up to 4–6 weeks to have the desired clinical effects¹⁵.

An important consideration for the therapeutic application of siRNA relates to its general safety, as well as to the safety profile associated with specific delivery technologies. General tolerability as well as specific toxicities (such as activation of complement, coagulation and cytokines) were evaluated for all monkeys in this study. We observed no treatment-related effects on the appearance or behaviour of animals treated with SNALP-siApoB-2 compared with saline-treated animals. There was no evidence for complement activation, delayed coagulation, pro-inflammatory cytokine production (Supplementary Table 1) or changes in haematology parameters (data not shown), toxicities that have been observed previously with treatments using related approaches^{16–19}. Across a systematic evaluation, the only detected change in primates treated with SNALP-siApoB-2 was a transient increase in liver enzymes in monkeys that received the high dose of SNALP-siApoB-2. The observed transaminosis peaked 48 h after treatment and was highly variable across individual animals. These effects, which were observed only at the highest dose of SNALP-siApoB-2, were completely reversible, with normalization by day 6 notwithstanding continued biological efficacy.

Our study highlights the potential for therapeutic gene silencing using systemic RNAi in non-human primates. A single, low dose of *APOB*-specific siRNA resulted in rapid and lasting RNAi-mediated gene silencing, with associated and profound phenotypic changes. The study was limited by the premature termination of the protocol

after 11 days, which prevented full evaluation of the time course for RNAi-mediated effects. Although further optimization of treatment regimen and safety profile characterization may be required, our data suggest that systemic delivery of siRNAs for targeting hepatocyte-specific genes in a higher species is possible. Furthermore, the rapid and long-lasting silencing of *APOB* using RNAi may represent a new strategy for reducing LDL-cholesterol in several relevant clinical settings.

METHODS

Additional details of the methods used are provided in the Supplementary Information.

siRNA formulation. The SNALP formulation contained the lipids 3-*N*-[(ω-methoxypoly(ethylene glycol)₂₀₀₀ carbamoyl]-1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleoyloxy-*N,N*-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar per cent ratio.

In vivo experiments. Saline and siRNA preparations were administered by tail vein injection under normal pressure and low volume (0.01 ml g⁻¹) for all rodent experiments. Cynomolgus monkeys ($n = 6$ per group) received either 2 ml kg⁻¹ phosphate buffered saline or 1 or 2.5 mg kg⁻¹ SNALP-siApoB-2 at a dose volume of 1.25 ml kg⁻¹ as bolus i.v. injections via the saphenous vein. For mRNA measurements, three liver biopsies per lobe were collected 48 h ($n = 4$) or 264 h ($n = 2$) after siRNA administration.

Bioanalytical methods. The QuantiGene assay (Genospectra) was used to quantify reduction in *APOB* mRNA levels relative to the housekeeping gene *GAPDH* in lysates prepared from mouse liver or cynomolgus monkey liver and jejunum as previously described⁶ but with minor variations. Mouse⁶ and cynomolgus monkey ApoB-100 protein levels were quantified by enzyme-linked immunosorbent assay (ELISA). LDL and HDL lipoprotein content were determined for plasma samples (250 μl) as described previously⁶.

Statistical analysis. *P*-values were calculated for comparison of SNALP-siApoB-2-treated animals with saline-treated animals using analysis of variance (ANOVA, two-factor without replication) with an alpha value of 0.05. *P*-values less than 0.05 were considered significant.

Received 12 January; accepted 6 March 2006.

Published online 26 March 2006.

- Novina, C. D. & Sharp, P. A. The RNAi revolution. *Nature* **430**, 161–164 (2004).
- Shankar, P., Manjunath, N. & Lieberman, J. The prospect of silencing disease using RNA interference. *J. Am. Med. Assoc.* **293**, 1367–1373 (2005).
- Thakker, D. R. *et al.* Neurochemical and behavioral consequences of widespread gene knockdown in the adult mouse brain by using nonviral RNA interference. *Proc. Natl Acad. Sci. USA* **101**, 17270–17275 (2004).
- Bitko, V., Musiyenko, A., Shulyayeva, O. & Barik, S. Inhibition of respiratory viruses by nasally administered siRNA. *Nature Med.* **11**, 50–55 (2005).
- Palliser, D. *et al.* An siRNA-based microbicide protects mice from lethal herpes simplex virus 2 infection. *Nature* **439**, 89–94 (2006).
- Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**, 173–178 (2004).
- Song, E. *et al.* Antibody mediated *in vivo* delivery of small interfering RNAs via cell-surface receptors. *Nature Biotechnol.* **23**, 709–717 (2005).
- Morrissey, D. V. *et al.* Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nature Biotechnol.* **23**, 1002–1007 (2005).
- Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).
- Cannon, C. P. *et al.* Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N. Engl. J. Med.* **350**, 1495–1504 (2004).
- Ridker, P. M. *et al.* C-reactive protein levels and outcomes after statin therapy. *N. Engl. J. Med.* **352**, 20–28 (2005).
- Crooke, R. M. *et al.* An apolipoprotein B antisense oligonucleotide lowers LDL cholesterol in hyperlipidemic mice without causing hepatic steatosis. *J. Lipid Res.* **46**, 872–884 (2005).
- Song, E. *et al.* Sustained small interfering RNA-mediated human immunodeficiency virus type 1 inhibition in primary macrophages. *J. Virol.* **77**, 7174–7181 (2003).
- Bartlett, D. W. & Davis, M. E. Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. *Nucleic Acids Res.* **34**, 322–333 (2006).
- Lennernas, H. & Fager, G. Pharmacodynamics and pharmacokinetics of the HMG-CoA reductase inhibitors. Similarities and differences. *Clin. Pharmacokinet.* **32**, 403–425 (1997).
- Levin, A. A. A review of the issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides. *Biochim. Biophys. Acta* **1489**, 69–84 (1999).

17. Chonn, A., Cullis, P. R. & Devine, D. V. The role of surface charge in the activation of the classical and alternative pathways of complement by liposomes. *J. Immunol.* **146**, 4234–4241 (1991).
18. Hornung, V. *et al.* Sequence-specific potent induction of IFN- α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nature Med.* **11**, 263–270 (2005).
19. Judge, A. D. *et al.* Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nature Biotechnol.* **23**, 457–462 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to P. Sharp, J. Maraganore and N. Mahanthappa for their assistance and support in this study. We would also like to thank W. J. Schneider, J. Frohlich, M. Hayden and J. E. Vance for

discussions. We acknowledge the technical assistance of C. Woppmann and A. Wetzel, and thank V. Kesavan and G. Wang for preparation of the cholesterol-conjugated siRNA used in this study. Finally, we thank S. Young for providing anti-ApoB antibodies. This work was supported by grants from the National Science and Engineering Research Council of Canada (to A.J.W. and M.N.F.).

Author Contributions This work represents the outcome of a collaboration between scientists at Alnylam Pharmaceuticals and Protiva Biotherapeutics Inc.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.S.Z. (tzimmermann@alnylam.com) or I.M. (ian@protivabio.com).

The depolymerizing kinesin MCAK uses lattice diffusion to rapidly target microtubule ends

Jonne Helenius^{1*}, Gary Brouhard^{1*}, Yannis Kalaidzidis^{1,2}, Stefan Diez¹ & Jonathon Howard¹

The microtubule cytoskeleton is a dynamic structure in which the lengths of the microtubules are tightly regulated. One regulatory mechanism is the depolymerization of microtubules by motor proteins in the kinesin-13 family¹. These proteins are crucial for the control of microtubule length in cell division^{2–4}, neuronal development⁵ and interphase microtubule dynamics^{6,7}. The mechanism by which kinesin-13 proteins depolymerize microtubules is poorly understood. A central question is how these proteins target to microtubule ends at rates exceeding those of standard enzyme–substrate kinetics⁸. To address this question we developed a single-molecule microscopy assay for MCAK, the founding member of the kinesin-13 family⁹. Here we show that MCAK moves along the microtubule lattice in a one-dimensional (1D) random walk. MCAK–microtubule interactions were

transient: the average MCAK molecule diffused for 0.83 s with a diffusion coefficient of $0.38 \mu\text{m}^2 \text{s}^{-1}$. Although the catalytic depolymerization by MCAK requires the hydrolysis of ATP, we found that the diffusion did not. The transient transition from three-dimensional diffusion to 1D diffusion corresponds to a “reduction in dimensionality”¹⁰ that has been proposed as the search strategy by which DNA enzymes find specific binding sites¹¹. We show that MCAK uses this strategy to target to both microtubule ends more rapidly than direct binding from solution.

Kinesin-13 motor proteins act at microtubule ends, where they are thought to force protofilaments into a curved conformation^{12,13}, which is a likely structural intermediate in the depolymerization process¹⁴. Classically, kinesin motor proteins reach microtubule ends by ATP-dependent translocation along microtubules. However,

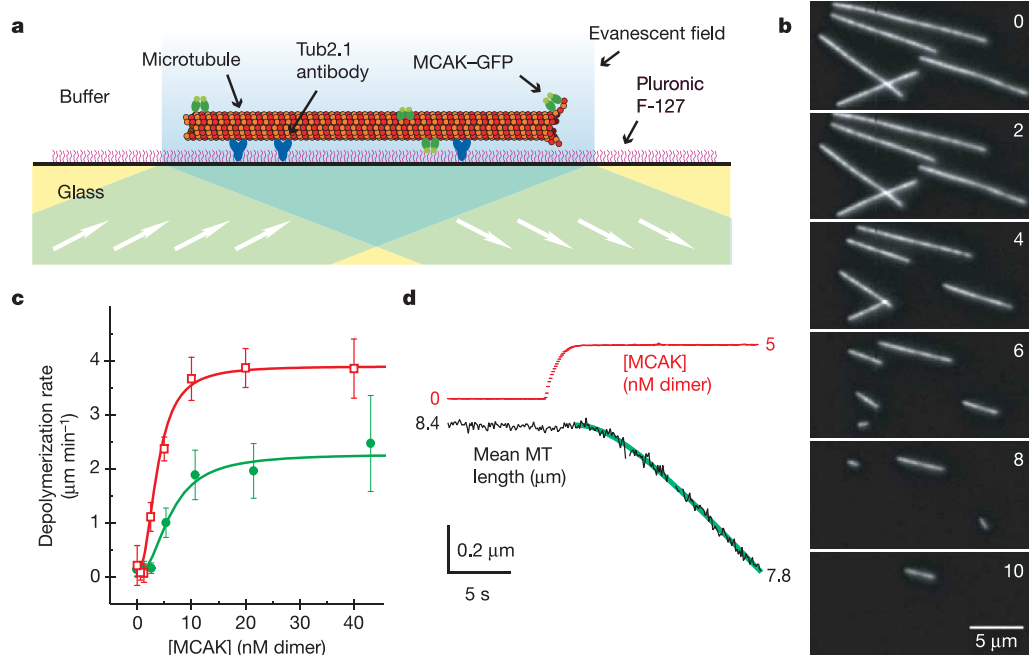


Figure 1 | MCAK-dependent microtubule depolymerization. **a**, Diagram of the *in vitro* assay depicting a microtubule (red) immobilized above the glass surface by anti-tubulin antibodies (dark blue). Excitation by total internal reflection allows the detection of single molecules (namely MCAK-GFP in green) in the evanescent field (shown in blue). **b**, Epifluorescence images of immobilized microtubules at the times shown in minutes. MCAK dimer (8 nM) was added at $t = 2$ min. **c**, Plot of microtubule depolymerization rate against MCAK concentration. Error

bars are s.d. Data fitted to Hill equations (lines plotted) yielded $K_m = 3.9$ nM and $K_m = 6.1$ nM for MCAK and MCAK-GFP, and $n = 2.4$ and $n = 2.2$, respectively. Red squares, MCAK-His₆; green circles, MCAK-His₆-EGFP. **d**, Shortening of four microtubules from a mean length of 8.4 μm to 7.8 μm (black line) after the addition of 5 nM MCAK (red line). The depolymerization rate approached steady state with a time constant of 3.8 s (green fitted line).

¹Max Planck Institute of Molecular Cell Biology and Genetics, Dresden 01307, Germany. ²A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119899, Russia.

*These authors contributed equally to this work.

directed motion has not been shown for kinesin-13 members⁸. To address the question of how kinesin-13 proteins reach microtubule ends, we expressed and purified two functional full-length versions of human MCAK: MCAK and green fluorescent protein (GFP)-tagged MCAK. We then developed a single-molecule microscopy assay for depolymerization (Fig. 1a) in which microtubules were immobilized on coverslips by means of surface-adsorbed anti-tubulin antibodies. Individual rhodamine-labelled microtubules and single MCAK–GFP molecules were revealed by epifluorescence and total-internal-reflection fluorescence (TIRF) illumination, respectively.

MCAK rapidly depolymerized GMP-CPP-stabilized microtubules in the depolymerization assay (Fig. 1b, and Supplementary movie 1). GMP-CPP (a slowly hydrolysable GTP analogue) was used to mimic the GTP found in the endcap of growing microtubules¹⁵. Depolymerization was ATP dependent and occurred at both microtubule ends. The depolymerization rate increased from its basal rate of $0.02 \mu\text{m min}^{-1}$ to a peak rate of $3.9 \mu\text{m min}^{-1}$ at 10 nM MCAK dimer (Fig. 1c). Thus, MCAK accelerated depolymerization up to 200-fold, with the maximum rate corresponding to about 50 tubulin dimers removed per second per microtubule end.

MCAK reached the microtubule ends very quickly. Microtubule lengths were measured after rapid infusion of MCAK into the observation chamber (Fig. 1d). On addition of 5 nM MCAK dimer, the depolymerization rate approached steady state with a rate constant, k , of 0.26 s^{-1} . This rate constant is related to the association rate to protofilament ends ($k_a = k[M]^{-1}$ when the MCAK concentration, $[M]$, is much greater than the Michaelis constant, K_m), and k_a must therefore be on the order of $50 \mu\text{M}^{-1} \text{ s}^{-1}$. Five experiments at either 5 or 10 nM MCAK gave an on-rate of $51 \pm 12 \mu\text{M}^{-1} \text{ s}^{-1}$ (results are means $\pm$ s.e.m. unless otherwise indicated). This confirms previous stopped-flow experiments⁸. Given that each microtubule has 14 protofilaments, we infer an association rate of about $700 \mu\text{M}^{-1} \text{ s}^{-1}$ to empty microtubule ends. These rates are significantly higher than expected for protein–protein association by three-dimensional diffusion in solution¹⁶; for example, the association of tubulin dimers onto the ends of growing microtubules occurs with an association rate of $2\text{--}5 \mu\text{M}^{-1} \text{ s}^{-1}$ (refs 16, 17).

These kinetic data suggest that the targeting of MCAK to the microtubule ends is somehow facilitated. To investigate this we performed experiments at MCAK–GFP concentrations that allowed single molecules to be observed (see Supplementary Information 1 for an analysis showing that single MCAK–GFP dimers were observed). Unlike other kinesins, which undergo directed movement on the microtubule lattice, we found that single MCAK–GFP molecules performed a random walk on the lattice during their transient interaction with a microtubule (Fig. 2a, and Supplementary movie 2). Using an in-house software package¹⁸, we tracked 1,147 MCAK–microtubule interactions that lasted longer than 0.4 s and calculated the diffusion coefficient, D , of MCAK–GFP to be $0.38 \pm 0.01 \mu\text{m}^2 \text{ s}^{-1}$ (Fig. 2b, Methods, and Supplementary Information 1). The interaction times were distributed exponentially (Fig. 2c), with a mean lifetime, $\langle t \rangle$, of $0.83 \pm 0.05 \text{ s}$ when corrected for photobleaching (Supplementary Information 1). At no time did we observe a directional bias in the diffusion (Supplementary Information 1).

The diffusion coefficient is high. Viewing diffusion as a random walk, the diffusion coefficient is related to the step size (δ) and the time per step (τ) by $D = \delta^2/2\tau$ (ref. 17). If $\delta = 8 \text{ nm}$ (corresponding to stepping between tubulin dimers along a protofilament), then $\tau = 0.084 \text{ ms}$, implying that MCAK takes 12,000 diffusive steps per second. This is more than 100-fold faster than kinesin-1 steps directionally along the microtubule lattice¹⁹.

To what extent does 1D diffusion increase the rate at which MCAK finds the ends of a microtubule? We calculated the average length scanned by MCAK during a diffusive interaction with a microtubule to be $0.79 \mu\text{m}$, calculated as $\sqrt{(2D\langle t \rangle)}$. This length of microtubule adjacent to the end acts as an ‘antenna’ for MCAK molecules (see

Supplementary Information 2). To estimate the flux of MCAK to the microtubule end, we write²⁰

$$\frac{\partial c}{\partial t} = k_{\text{on}}C_m - k_{\text{off}}c + D\frac{\partial^2 c}{\partial x^2}$$

where $c(x, t)$ is the concentration of MCAK on the microtubule lattice at distance x from the microtubule end and time t , k_{on} is the attachment rate to the lattice, C_m is the MCAK concentration in solution, and k_{off} is the dissociation rate from the lattice. This equation excludes the contribution of the depolymerization rate, which is negligible. The steady-state solution of this equation is given by $c(x) = c_\infty(1 - e^{-x/x_0})$, where c_∞ is the concentration on the microtubule lattice far from the end, in MCAK dimers per μm , and $x_0 = \sqrt{(D/k_{\text{off}})}$. The flux to the microtubule end is $J = -D\partial c/\partial x = -Dc_\infty/x_0$ (see Supplementary Information 3). c_∞ was measured for MCAK–GFP concentrations of 0.3–3 nM by direct counting of molecules. This allowed us to calculate the attachment rate to the microtubule lattice, $k_{\text{on}} = 0.64 \pm 0.13 \text{ nM}^{-1} \text{ s}^{-1} \mu\text{m}^{-1}$ ($n = 3$). With this value of k_{on} , we calculated a flux to the end of the microtubule of 2.2 MCAK–GFP dimers per second at 6 nM, corresponding to an association rate to protofilament ends of $k_a = 26 \mu\text{M}^{-1} \text{ s}^{-1}$. Thus, the diffusion of MCAK accounts theoretically for the very high association rate measured in the initial-rate experiments (Fig. 1d).

Diffusive scanning followed by end capture was directly observed, as shown in Fig. 3a (yellow arrowhead). If an MCAK–GFP molecule lands within $0.25 \mu\text{m}$ of the microtubule end, it is likely to diffuse to the end within our 100-ms frame acquisition time; this accounts for end-binding events not preceded by observable diffusion

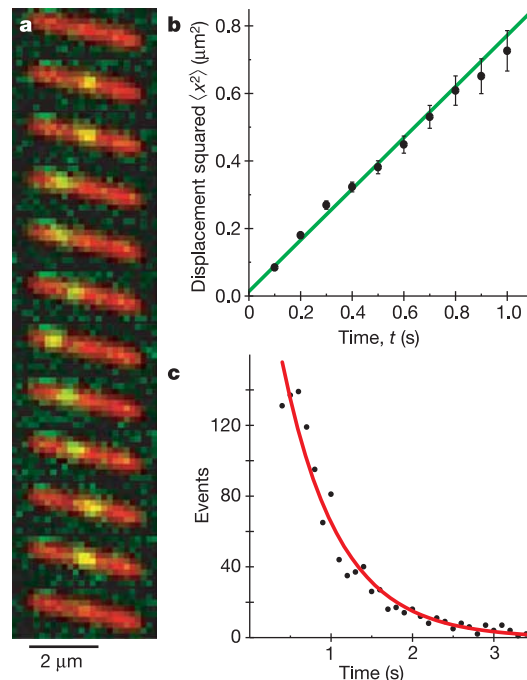


Figure 2 | MCAK–GFP diffusion along microtubules. **a**, Sequential frames of an MCAK–GFP (green) video. Twelve frames of a continuous TIRF-FITC recording (100-ms frames) were overlaid on one TRITC epifluorescence image of the microtubule (red). **b**, The mean-squared displacement of MCAK is plotted against the time interval over which it is measured. A linear curve fitted to the initial second yields a diffusion coefficient, D , of $0.38 \mu\text{m}^2 \text{ s}^{-1}$; $\langle x^2 \rangle = 2Dt$. Error bars represent the s.e.m. of the squared displacement values. **c**, Histogram of durations of MCAK–GFP–microtubule interactions. An exponential curve fitted to the histogram (red) and corrected for photobleaching yields a mean lifetime of the interactions, $\langle t \rangle$, of 0.83 s.

(Fig. 3a, white arrowhead). Direct end binding from solution also occurs.

Why might MCAK use 1D diffusion instead of a directed walk to get to microtubule ends? One advantage is that, unlike directed motion, diffusion allows MCAK to target both ends of the microtubule, which is a significant feature of kinesin-13 localization *in vivo*^{2,4}. A second advantage becomes clear when the diffusive movement of MCAK (with a diffusion coefficient of $0.38 \mu\text{m}^2 \text{s}^{-1}$) is compared with the directed movement of kinesin-1 (with a velocity of $0.8 \mu\text{m s}^{-1}$)¹⁹: MCAK covers shorter distances more rapidly (Fig. 3b). For any distance shorter than $1 \mu\text{m}$, MCAK will outpace kinesin-1 in a race to the microtubule end.

We found evidence that the rapid diffusion of MCAK to the microtubule end does indeed accelerate the depolymerization reaction. First, increasing the KCl concentration from 75 mM to 125 mM greatly decreased the lifetime of MCAK on the microtubule lattice and decreased the depolymerization rate to baseline levels (Fig. 3c and Supplementary Information 4). Because a high salt concentration may disrupt direct end association as well as lattice diffusion, we did a second experiment in which the negatively charged carboxy termini of tubulin, known as the 'E-hook'²¹, was removed by digestion with subtilisin. This prevented MCAK from diffusing on the lattice (Fig. 3d, e) and significantly decreased the depolymerization rate to $26 \pm 3\%$ of control microtubules ($P < 0.1\%$, Welch's *t*-test), as shown previously with microtubule sedimentation

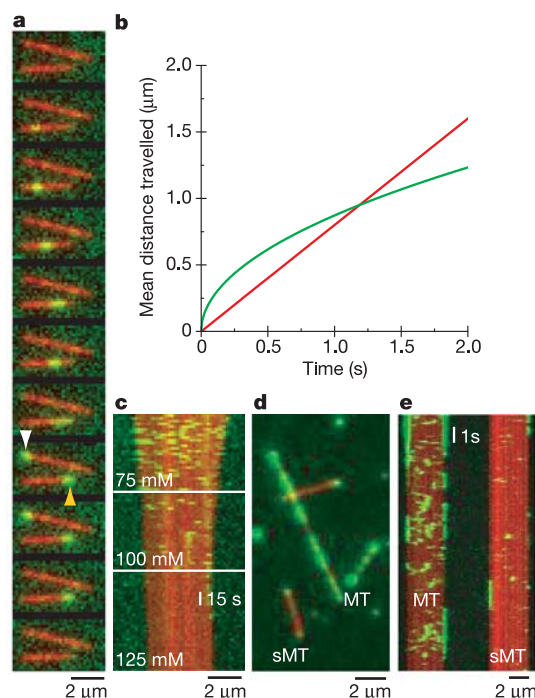


Figure 3 | MCAK targeting of microtubule ends by means of diffusion.

a, Two MCAK-GFP molecules (green) interacting with separate microtubules (red) in 1 mM ATP (100-ms frames). Arrowheads indicate MCAK-microtubule end-binding events. **b**, Plot of average distance travelled over time by diffusing MCAK-GFP (green, $D = 0.38 \mu\text{m}^2 \text{s}^{-1}$; $\langle x \rangle = \sqrt{2Dt}$) and walking kinesin-1 (red, velocity $v = 0.8 \mu\text{m s}^{-1}$; $\langle x \rangle = vt$). **c**, Kymograph showing the effect of KCl concentration on the MCAK interaction with a single microtubule. Microtubules are depicted along the horizontal axis while time changes along the vertical axis. The MCAK and MCAK-GFP concentrations were 2.5 and 0.25 nM, respectively. **d**, Dual-colour image of normal microtubules (MT) and subtilisin-digested microtubules (sMT) in the presence of 3 nM MCAK-GFP (green). **e**, Kymographs showing diffusion on a normal microtubule (MT) and only brief binding/unbinding events on a subtilisin-digested microtubule (sMT) at about 1 nM MCAK-GFP.

assays^{12,13}. Direct end binding still occurred, and nearly maximum depolymerization rates were observed when using very high concentrations of MCAK (Supplementary Information 4). Therefore the decrease in depolymerization rate for subtilisin microtubules is best explained by a decrease in end-targeting caused by the lack of E-hook-mediated diffusion.

The electrostatic partner for the negatively charged E-hook is probably the positively charged 'neck' domain²², which has been shown to be necessary for MCAK depolymerization activity *in vivo*²³. Complementing earlier work showing that electrostatic interactions of positively charged regions of kinesin-1 (refs 24, 25) and kinesin-3 (ref. 26) with the E-hook enhance processivity of directed motility, our results indicate that these interactions are important for diffusive motility as well. Taken together, these data connect diffusion and depolymerization: electrostatic interactions between MCAK's neck domain and the E-hook of tubulin give rise to diffusion and accelerated depolymerization.

ATP is hydrolysed while MCAK interacts with the lattice of microtubules⁸. It is clear that each individual diffusive step does not require ATP hydrolysis, because MCAK's lattice-stimulated ATPase activity was previously measured to be only 1s^{-1} (ref. 8), much lower than the rate of diffusive stepping ($12,000 \text{s}^{-1}$). The mean interaction time of 0.83 s indicates that one diffusive interaction corresponds to one ATP hydrolysis cycle.

To determine whether diffusion depended on MCAK's nucleotide state, we performed single-molecule diffusion experiments with ADP, AMP-PNP (a non-hydrolysable ATP analogue) or apyrase (to digest all residual ATP and ADP) to mimic the ADP, ATP or

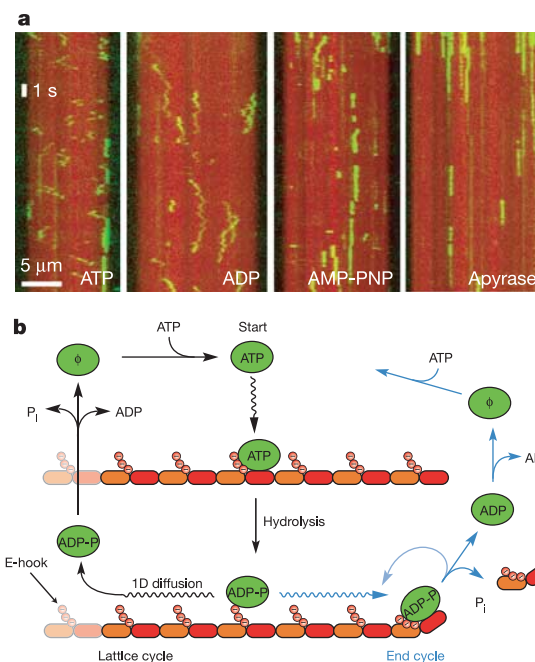


Figure 4 | MCAK nucleotide states and diffusion model. **a**, Kymographs depicting the motion of MCAK-GFP (green) along microtubules (red) in the presence of ATP (1 mM), ADP (1 mM), AMP-PNP (1 mM) and apyrase. The MCAK-GFP concentration was 0.6 nM except with apyrase, when it was 20 pM. **b**, MCAK-microtubule interaction model. Binding of soluble MCAK-ATP (green) induces rapid hydrolysis, accounting for the lattice-stimulated ATPase activity⁸. Next, MCAK-ADP-P_i diffuses along the lattice by means of the E-hook of tubulin. It may disassociate, 'unproductively' completing the hydrolysis cycle (left, black arrows). Alternatively, it may reach a microtubule end, where the removal of tubulin dimers is coupled to P_i release (right, blue arrows), accounting for the end-stimulated ATPase activity⁸. MCAK may detach from the end (straight blue arrow) or, if processive, may stay attached to remove additional dimers (light blue arrow).

nucleotide-free states, respectively. In 1 mM ADP, MCAK–GFP bound microtubules and moved in a diffusive manner ($D = 0.22 \pm 0.01 \mu\text{m}^2 \text{s}^{-1}$, $\langle t \rangle = 1.75 \pm 0.06 \text{ s}$, $n = 1,076$) proving that diffusion occurred without energy derived from the hydrolysis of ATP (Fig. 4a, and Supplementary movie 3). Similar diffusion coefficients were measured in 0.1 and 10 mM ADP. The statistically significant difference in diffusion parameters (D and $\langle t \rangle$) in ADP compared with ATP implies that, in the presence of ATP, MCAK does not primarily diffuse in the ADP state. MCAK–AMP–PNP could bind to microtubules but only rarely showed diffusive steps ($D = 0.014 \mu\text{m}^2 \text{s}^{-1}$, $\langle t \rangle = 3.5 \text{ s}$). When apyrase was added to remove both ATP and ADP from solution enzymatically, nucleotide-free MCAK–GFP went into a rigor state with greatly reduced diffusion ($D = 0.003 \mu\text{m}^2 \text{s}^{-1}$, $\langle t \rangle = 3.9 \text{ s}$).

In the presence of ATP, MCAK must diffuse in one or more of the four possible nucleotide states: nucleotide-free, ATP, ADP and ADP– P_i . As MCAK binds tightly in the presence of AMP–PNP and apyrase, both the ATP and nucleotide-free states are excluded. Because the diffusion observed in ADP is statistically distinct from that observed in ATP, MCAK apparently diffuses in the ADP– P_i state. To induce an ADP– P_i -like state, we performed a single-molecule diffusion experiment in ADP·AlF₃, ADP·BeF₃ and ADP·vanadate. Diffusion was observed in all three cases (Supplementary Information 5). The results imply that MCAK normally, in the presence of ATP, diffuses in the ADP– P_i state.

In what nucleotide state is MCAK when it first encounters the microtubule? To determine the nucleotide state of soluble MCAK, we measured the amounts of radiolabelled α - and γ -phosphates of ATP bound to MCAK and kinesin-1 free in solution without microtubules (Supplementary Information 5). As shown previously²⁷, no γ - P_i was bound to kinesin (γ - P_i/α - $\text{P}_i = -11 \pm 8\%$ in two experiments). However, equal amounts of α - P_i and γ - P_i remained bound to MCAK (γ - P_i/α - $\text{P}_i = 116 \pm 11\%$ in three experiments). This shows that MCAK exists in the ATP or ADP– P_i state in solution, unlike kinesin-1, and indicates that the functional differences between kinesin-1 and kinesin-13 family proteins are achieved through differences in hydrolysis mechanisms. The nucleotide data were used to formulate a MCAK–microtubule interaction model that accounts for both the microtubule lattice-stimulated and end-stimulated ATPase activity of MCAK (Fig. 4b).

After arriving at the end, each MCAK seems to remove several dimers. Comparison of the flux to the microtubule end calculated theoretically (2.2 MCAK dimers per second at 6 nM) with the off-rate of tubulin dimers (14 s^{-1} at 6 nM MCAK–GFP) indicates that each MCAK removes about seven tubulin dimers for each end-targeting event. On average, single MCAK end-binding events were observed to last $1.9 \pm 0.2 \text{ s}$ ($n = 239$, corrected for photobleaching). At saturating concentration, MCAK removes tubulin dimers from protofilaments at a rate of two dimers per second, so an MCAK that resides for 1.9 s at an end would remove about four tubulin dimers. Taken together, these results indicate that MCAK might act processively at the microtubule end.

By diffusing along microtubules instead of walking, MCAK rapidly targets both microtubule ends. A comparable process is the rapid targeting of DNA restriction enzymes to their restriction sites, which also exceeds the three-dimensional diffusion limit²⁸. These DNA enzymes are thought to target their restriction sites by a 1D diffusional scan of DNA segments^{29,30}, but this search mechanism has not been observed at the single-molecule level. Here we have characterized the 1D diffusion of MCAK on microtubules and shown the implications of this diffusion for end targeting. The single-molecule data presented here support the ‘reduction in dimensionality’ hypothesis.

The MCAK neck domain data²³ suggest that 1D diffusion operates *in vivo* to target MCAK to microtubule ends. Once at the ends, MCAK and its homologues interact with other proteins, such as the plus-end-binding proteins XMAP215 and EB1 (refs 6, 7), to regulate

microtubule dynamics. The *in vitro* single-molecule approach taken here could be useful to study the dynamics of this molecular machinery.

METHODS

Proteins. Human MCAK–His₆ and human MCAK–His₆ tagged with enhanced GFP were expressed in *Spodoptera frugiperda* (Sf9) cells (BAC-TO-BAC expression system; Invitrogen) and purified by cation-exchange, metal-chelating, and desalting or gel-filtration chromatography. Most experiments with MCAK–GFP were performed with freshly purified protein. A filter-based radiometric ATP binding assay⁸ with [γ -³²P]ATP and [α -³²P]ATP was used to determine the concentration of active MCAK and the nucleotide state of proteins in solution. Details are given in Supplementary Information 5 and 6. Reagents were purchased from Sigma unless indicated otherwise. Pig-brain tubulin was purified, rhodamine-labelled (TAMRA; Invitrogen) and polymerized with GMP–CPP (Jena Bioscience) as described previously⁸. Key experiments were repeated with Taxol-stabilized microtubules; no differences were observed. Digestion was performed with $10 \mu\text{g ml}^{-1}$ subtilisin, incubated for 20 min at 37 °C, and terminated with 2 mM phenyl-methylsulphonyl fluoride.

Imaging. Images were acquired with either a Roper Scientific MicroMAX:512BFT charge-coupled device camera or an Andor DV887 iXon camera with Zeiss Axiovert 200M microscopes and Zeiss 100 $\times$ /1.45 α Plan-FLUAR objectives. The microscopes were outfitted with a dual-port TIRF condenser (Till Photonics) or a prototype VisiTIRF condenser (Visitron Systems). Fluorescein isothiocyanate (FITC) and tetramethylrhodamine β -isothiocyanate (TRITC) filter sets (Chroma Technology Corp.) were used to image GFP (TIRF) and TAMRA (epifluorescence) fluorophores, respectively. The standard exposure time was 100 ms.

Depolymerization assay. Microscope chambers were constructed with 18 mm $\times$ 18 mm and 22 mm $\times$ 22 mm coverslips separated by double-sided tape (Scotch 3M) to create channels 0.1 mm thick, 3 mm wide and 18 mm long. Glass coverslips (no. 1.5; Corning) were cleaned in Piranha solution ($\text{H}_2\text{O}_2/\text{H}_2\text{SO}_4$, 3:5) before silanization in 0.05% dichlorodimethylsilane in trichloroethylene. Detailed methods are given in Supplementary Information 6. To immobilize microtubules, channels were incubated with 0.2% Tub 2.1 antibody in BRB80 for 5 min, followed by 5 min with 1% Pluronic F-127 in BRB80, and finally GMP–CPP microtubules in BRB80 for 15 min. Channels were rinsed with BRB20 (20 mM PIPES/KOH pH 6.9, 1 mM MgCl_2 , 1 mM EGTA) before addition of the imaging solution (BRB20 supplemented with 75 mM KCl, 0.1 mg ml^{−1} BSA, 1 mM ATP or other nucleotides, 1% 2-mercaptoethanol, 40 mM glucose, 40 $\mu\text{g ml}^{-1}$ glucose oxidase, 16 $\mu\text{g ml}^{-1}$ catalase, and MCAK).

Imaging analysis. The Motion Tracking software package, written in the Pluk development environment, was used to locate and track MCAK–GFP molecules¹⁸. The validity of each track was confirmed by visual inspection. The Pluk-derived diffusion coefficient was cross-checked against, first, trajectories generated by hand with kymographs and a custom MATLAB-based two-dimensional gaussian peak-fitting tool, and second, an analysis of displacement distributions that does not require manual inspection. Details are given in Supplementary Information 1. For each condition examined, at least three separate experiments were performed, and multiple films of each experiment were analysed, yielding no less than 100 molecules.

Received 16 February; accepted 17 March 2006.

- Desai, A., Verma, S., Mitchison, T. J. & Walczak, C. E. Kin I kinesins are microtubule-destabilizing enzymes. *Cell* **96**, 69–78 (1999).
- Maney, T., Hunter, A. W., Wagenbach, M. & Wordeman, L. Mitotic centromere-associated kinesin is important for anaphase chromosome segregation. *J. Cell Biol.* **142**, 787–801 (1998).
- Rogers, G. C. *et al.* Two mitotic kinesins cooperate to drive sister chromatid separation during anaphase. *Nature* **427**, 364–370 (2004).
- Walczak, C. E., Mitchison, T. J. & Desai, A. XKCM1: a *Xenopus* kinesin-related protein that regulates microtubule dynamics during mitotic spindle assembly. *Cell* **84**, 37–47 (1996).
- Homma, N. *et al.* Kinesin superfamily protein 2A (KIF2A) functions in suppression of collateral branch extension. *Cell* **114**, 229–239 (2003).
- Tournebise, R. *et al.* Control of microtubule dynamics by the antagonistic activities of XMAP215 and XKCM1 in *Xenopus* egg extracts. *Nature Cell Biol.* **2**, 13–19 (2000).
- Mennella, V. *et al.* Functionally distinct kinesin-13 family members cooperate to regulate microtubule dynamics during interphase. *Nature Cell Biol.* **7**, 235–245 (2005).
- Hunter, A. W. *et al.* The kinesin-related protein MCAK is a microtubule depolymerase that forms an ATP-hydrolyzing complex at microtubule ends. *Mol. Cell* **11**, 445–457 (2003).

9. Wordeman, L. & Mitchison, T. J. Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis. *J. Cell Biol.* **128**, 95–104 (1995).
10. Adam, G. & Delbruck, M. in *Structural Chemistry of Molecular Biology* (eds Rich, A. & Davidson, N.) 198–215 (Freeman, San Francisco, 1968).
11. Richter, P. H. & Eigen, M. Diffusion controlled reaction rates in spheroidal geometry. Application to repressor–operator association and membrane bound enzymes. *Biophys. Chem.* **2**, 255–263 (1974).
12. Moores, C. A. *et al.* A mechanism for microtubule depolymerization by KinI kinesins. *Mol. Cell* **9**, 903–909 (2002).
13. Niederstrasser, H., Salehi-Had, H., Gan, E. C., Walczak, C. & Nogales, E. XKCM1 acts on a single protofilament and requires the C terminus of tubulin. *J. Mol. Biol.* **316**, 817–828 (2002).
14. Mandelkow, E. M., Mandelkow, E. & Milligan, R. A. Microtubule dynamics and microtubule caps: a time-resolved cryo-electron microscopy study. *J. Cell Biol.* **114**, 977–991 (1991).
15. Hyman, A. A., Salser, S., Drechsel, D. N., Unwin, N. & Mitchison, T. J. Role of GTP hydrolysis in microtubule dynamics: information from a slowly hydrolyzable analogue, GMPCPP. *Mol. Biol. Cell* **3**, 1155–1167 (1992).
16. Northrup, S. H. & Erickson, H. P. Kinetics of protein–protein association explained by Brownian dynamics computer simulation. *Proc. Natl Acad. Sci. USA* **89**, 3338–3342 (1992).
17. Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland, Massachusetts, 2001).
18. Kalaidzidis, Y. L., Gavrilov, A. V., Zaitsev, P. V., Kalaidzidis, A. L. & Korolev, E. V. PLUK—an environment for software development. *Program. Comput. Softw.* **23**, 206–211 (1997).
19. Howard, J., Hudspeth, A. J. & Vale, R. D. Movement of microtubules by single kinesin molecules. *Nature* **342**, 154–158 (1989).
20. Klein, G. A., Kruse, K., Cuniberti, G. & Jülicher, F. Filament depolymerization by motor molecules. *Phys. Rev. Lett.* **94**, 108102 (2005).
21. Paschal, B. M., Obar, R. A. & Vallee, R. B. Interaction of brain cytoplasmic dynein and MAP2 with a common sequence at the C terminus of tubulin. *Nature* **342**, 569–572 (1989).
22. Maney, T., Wagenbach, M. & Wordeman, L. Molecular dissection of the microtubule depolymerizing activity of mitotic centromere-associated kinesin. *J. Biol. Chem.* **276**, 34753–34758 (2001).
23. Ovechkina, Y., Wagenbach, M. & Wordeman, L. K-loop insertion restores microtubule depolymerizing activity of a ‘neckless’ MCAK mutant. *J. Cell Biol.* **159**, 557–562 (2002).
24. Thorn, K. S., Ubersax, J. A. & Vale, R. D. Engineering the processive run length of the kinesin motor. *J. Cell Biol.* **151**, 1093–1100 (2000).
25. Wang, Z. & Sheetz, M. P. The C-terminus of tubulin increases cytoplasmic dynein and kinesin processivity. *Biophys. J.* **78**, 1955–1964 (2000).
26. Nitta, R., Kikkawa, M., Okada, Y. & Hirokawa, N. KIF1A alternately uses two loops to bind microtubules. *Science* **305**, 678–683 (2004).
27. Hackney, D. D. Kinesin ATPase: Rate-limiting ADP release. *Proc. Natl Acad. Sci. USA* **85**, 6314–6318 (1988).
28. Riggs, A. D., Bourgeois, S. & Cohn, M. The lac repressor–operator interaction. 3. Kinetic studies. *J. Mol. Biol.* **53**, 401–417 (1970).
29. Winter, R. B., Berg, O. G. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry* **20**, 6961–6977 (1981).
30. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Hartmann for cloning and protein purification work; V. Varga for purification of unlabelled MCAK; E. Schäffer for the application of F-127; F. Ruhnnow for the two-dimensional gaussian peak-fitting tool; A. Hyman, F. Jülicher, E. Schäffer, I. Riedel, J. Stear, G. Klein, V. Varga, A. Hunt and H. T. Schek for review of the manuscript; and S. Wolfson for proofreading and editing. Andor Technologies lent us the Andor iXon camera, and the VisiTRF condenser was developed in collaboration with Visitron Imaging Systems, GmbH. This research was supported by the Max Planck Gesellschaft and the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H. (howard@mpi-cbg.de).

ERRATUM

doi:10.1038/nature04774

Significant primordial star formation at redshifts $z \approx 3-4$

Raul Jimenez & Zoltan Haiman

Nature 440, 501–504 (2006)

In this Letter, the received and accepted dates were incorrect. They should read: 'Received 6 September 2005; accepted 4 January 2006.'

CORRIGENDUM

doi:10.1038/nature04726

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, Michael S. Braverman, Yi-Ju Chen, Zhoutao Chen, Scott B. Dewell, Alex de Winter, James Drake, Lei Du, Joseph M. Fierro, Robin Forte, Xavier V. Gomes, Brian C. Godwin, Wen He, Scott Helgesen, Chun Heen Ho, Stephen K. Hutchison, Gerard P. Irzyk, Szilveszter C. Jando, Maria L. I. Alenquer, Thomas P. Jarvie, Kshama B. Jirage, Jong-Bum Kim, James R. Knight, Janna R. Lanza, John H. Leamon, William L. Lee, Steven M. Lefkowitz, Ming Lei, Jing Li, Kenton L. Lohman, Hong Lu, Vinod B. Makhijani, Keith E. McDade, Michael P. McKenna, Eugene W. Myers, Elizabeth Nickerson, John R. Nobile, Ramona Plant, Bernard P. Puc, Michael Reifler, Michael T. Ronan, George T. Roth, Gary J. Sarkis, Jan Fredrik Simons, John W. Simpson, Maithreya Srinivasan, Karrie R. Tartaro, Alexander Tomasz, Kari A. Vogt, Greg A. Volkmer, Shally H. Wang, Yong Wang, Michael P. Weiner, David A. Willoughby, Pengguang Yu, Richard F. Begley & Jonathan M. Rothberg

Nature 437, 376–380 (2005)

In this paper, the second name of Chun Heen Ho was misspelled as He. In the earlier Corrigendum (*Nature* 439, 502; 2006) the surnames of Stephen K. Hutchison and Brian C. Godwin were misspelled as Hutchinson and Goodwin. In addition, in the Supplementary Information of the original paper, there were errors in the sequences of the DNA capture and enrichment primers, and in the formulation of emulsion oil. The Supplementary Information was updated on 4 May 2006.

CORRIGENDUM

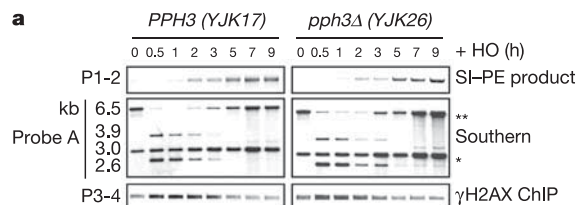
doi:10.1038/nature04772

A phosphatase complex that dephosphorylates γ H2AX regulates DNA damage checkpoint recovery

Michael-Christopher Keogh, Jung-Ae Kim, Michael Downey, Jeffrey Fillingham, Dipanjan Chowdhury, Jacob C. Harrison, Megumi Onishi, Nira Datta, Sarah Galicia, Andrew Emili, Judy Lieberman, Xueting Shen, Stephen Buratowski, James E. Haber, Daniel Durocher, Jack F. Greenblatt & Nevan J. Krogan

Nature 439, 497–501 (2006)

Figure 3a of this Letter contains an inadvertently duplicated panel set: those referring to '*pph3 Δ* (YJK26)' are identical to '*PPH3* (YJK17)'. The corrected panels are shown here. Note that the quantification shown in Fig. 3b refers to the correct panels. Our results and conclusions are unaffected by this oversight.



CORRIGENDUM

doi:10.1038/nature04773

A high-mobility electron gas at the LaAlO₃/SrTiO₃ heterointerface

A. Ohtomo & H. Y. Hwang

Nature 427, 423–426 (2004)

In Fig. 2a and b of this Letter, the y axis should be labelled in units of ohms per square, not milli-ohms per square. This error does not occur elsewhere in the paper, and does not change any of our results or conclusions.

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Thinking big

If consistency is the hobgoblin of little minds, as Ralph Emerson had it, then Europe's purveyors of career information are definitely broad-minded. European data on scientific careers are all over the place. The information is hard to understand because it differs from country to country, and varying definitions confound it further still.

But a workshop in Madrid last week aimed to cut through that inconsistency. 'Research careers for the 21st century' was sponsored by the Paris-based Organisation for Economic Co-operation and Development (OECD) and Spain's science and education ministry, and it offered some solid proposals for clarifying career options for young scientists in OECD countries.

In kicking off the workshop, Nobuo Tanaka, director for science, technology and industry at the OECD, said that future career trajectories may stretch beyond the standard path that most prospective young scientists are offered. Institutions should think "outside the traditional linear pathway", Tanaka said. Many still see academia, industry and government jobs as distinct paths, and such barriers offer a challenge, he added.

Another barrier is posed by the quality of career data currently on offer. Ulrich Teichler, of the University of Kassel's International Centre for Higher Education Research, jokes that it is an "absolute disaster" trying to compare career data for European scientists. He points out that the age at which students enter PhDs varies from country to country, as do definitions of when these degrees are complete and when a scientist is considered independent. Despite these difficulties, Teichler has drawn some solid conclusions about European scientists: only about 22% of doctorates go on to tenure-track careers and the average age at which academic scientists reach independent status is 40 to 42 years old.

These figures may vary according to discipline and country, but they are consistently worth communicating to young scientists early in their careers.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik

European Head Office, London
The Macmillan Building,
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
Natureevents: Matt Betteridge (4015)

UK/RoW/Ireland:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Claudia Paulsen Young (4973)
France/Switzerland/Belgium/Italy: Maria Ceraolo (4994)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Business Development Manager:
Amelie Pequignot (4974)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.

Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

US Head Office, New York
75 Varick Street,
9th Floor, New York,

NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746
Asia-Pacific Sales Manager: Ayako Watanabe
e-mail: a.watanabe@natureasia.com

Misconduct mayhem

naturejobs

**Survive
in
Science**
Volume 3
**Scientific
misconduct**

1. Don't stand by and let fraudulent research taint your career.



2. Verify with trusted colleagues and report to the proper channels



As a lab technician at the University of Vermont in Burlington, Walter DeNino had a gut feeling that something was amiss when his boss repeated a statistical analysis that DeNino had done and arrived at a different conclusion. DeNino followed his hunch and spent his evenings reconstructing the data set from patient medical records. He found that the original data did not match his supervisor's figures.

DeNino had uncovered one of the most serious cases of scientific misconduct reported in recent years. His boss, obesity expert Eric Poehlman, had committed scientific fraud for more than 12 years in numerous publications and grant proposals¹. Now debarred from receiving federal research funding for life, Poehlman must repay \$180,000 and is one of only two researchers ever charged in a US criminal court for misconduct.

"I felt that his behaviour had to be exposed and that he should be removed from science," says DeNino, now a researcher in another obesity lab group at Weill Medical College, Cornell University in New York. "But at the onset, I didn't imagine it would take five years and that criminal charges would be filed."

After an arduous and sometimes ugly investigation process, DeNino says the experience opened new research opportunities for him and spurred his desire to become a physician-researcher. But he admits he was lucky. As a whistleblower, he was very cautious in making an allegation and he was well-protected by the policies and actions of the University of Vermont.

Known cases of misconduct are rare, involving one or two scientists in 100. Chances are, however, that most scientists at some time will come across a lab member's notebook, a collaborator's work, a manuscript, grant proposal or thesis that seems suspect. Knowing the proper way to approach and investigate will help you handle such a sticky situation.

Lessons can be learned from the cases of three junior scientists who found misconduct by senior researchers, did or did not come forward, and the consequences of

Thinking about scientific misconduct before tangling with a real case will help you protect your own career and promote research integrity.

Kendall Powell investigates a few case studies.

their actions. Misconduct by a supervisor puts the junior scientist in a very vulnerable position. But with the proper tools and advice, a graduate student, postdoc or junior faculty member should be able to make an allegation without risking her or his career.

Corrupt or just sloppy?

It is vital to know what constitutes misconduct. The US Public Health Service defines it as "fabrication, falsification or plagiarism in proposing, performing or reviewing research, or in reporting research results"². It does not include honest mistakes or differences in interpretation. It must be shown to have been committed intentionally or recklessly and be a significant departure from accepted practices in a field. This 'FFP' definition is widely accepted by universities and other government agencies.

Some unethical or poor research methods fall short of being FFP. They include keeping inadequate records, putting speculation forth as fact, using confidential information and dropping data points from statistical analyses. These are much more common than FFP, with close to a third of scientists engaging in such questionable practices³.

"We as a community need to work on research integrity a little bit more," says Julio Turrens, associate dean and biochemist at the University of South Alabama in Mobile. He teaches courses on research integrity in the university's school of medicine.

Although misconduct cases are few, their effects on the researchers involved, the field and ultimately on society can be far-reaching, says Turrens. He gives students examples of fraudulent studies that changed environmental regulations or had breast-cancer patients clamouring for an ineffective treatment. He warns students that if they make an allegation of misconduct to him, as a dean he is bound to investigate it. Most institutions have a protocol for investigating scientific fraud, and Turrens advises scientists to consult two or

three sources of confidential advice before passing “the point of no return”.

DeNino was successful because he knew where that point was in his university's investigation policies. He set up an anonymous e-mail account so that he could get advice and information from university offices before stepping forward formally.

Turens suggests doing your homework to find the university office that handles investigations — often called the ‘office of research compliance’ or ‘research integrity’ — and figuring out how to report something directly to that office. “Often that puts the allegation out of the loop of the department or school, then there is no allegiance between the person you are talking to and the person you are complaining about,” he says.

DeNino got trusted colleagues to double-check his suspicions, which research-integrity experts strongly recommend. Margaret Dale, dean for faculty and research integrity at Harvard Medical School, advises finding someone trustworthy to test that you really understand what's happening. “Ask in a non-threatening way, ‘I see this happening, what do you think?’,” she says. Then if you do proceed “there's another person who knows what's going on and safety in numbers”.

Kent*, a postdoc in a US West Coast institution, says his allegation against his former adviser might have had a better outcome had he found senior scientists to back him. “It is very difficult to find faculty members to take your side in public,” he says. Although some supported his findings, they were unwilling to confront his adviser. Kent's situation highlights the vulnerability of postdocs who come into an institution with a one-to-one relationship with an adviser. It easily becomes a case of one's word against the other's.

Kent discovered his adviser had taken parts of a grant proposal Kent had written based on his graduate work and used them, verbatim, for another postdoc's fellowship application, giving no credit to Kent's previous lab and work. Kent filed a formal grievance with the vice-chancellor for research. But because the original proposal was submitted under his adviser's name, and because Kent had no proof of intent to steal credit, the grievance resulted in a rather unsatisfying letter of apology — and, Kent believes, retaliation.

“In this situation, you are pretty powerless,” Kent

Julio Turens: “We need to work on research integrity.”



says. “Be very careful about writing a grant application submitted under someone else's name. The investigator has too much power in that situation.” He advises getting a written contract that spells out your expected contribution to research and the

time frame you are expected to work in.

Integrity officers stress that junior scientists must find senior scientists or officials who will support them if they bring forward an allegation. Dale says postdocs especially should find out what resources are available to them — postdoc office, ombudsman office, and so on — before moving ahead with a complaint.

Dale notes that, with most allegations arising within a lab group, it is nearly impossible to keep the accuser's identity confidential from the accused. Junior scientists must understand the risks they will face if they become known as a whistleblower and the consequences their actions may have, such as shutting down a laboratory.

Stakes too high

Emma* weighed up the risks against some dubious practices in her graduate thesis laboratory at a private research institute. Although her adviser verbally presented data at meetings that she knew to be false, she did not speak out. She and lab colleagues discussed the matter, but felt the stakes were too high, as some would lose jobs or prospects if the lab was disgraced.

Such conflicts are normal, says Chris Pascal, director of the Office of Research Integrity, which monitors investigations of misconduct in research supported by US Public Health Service funds. “Unless the misconduct is so serious that a human life is at stake, I would never say that someone has to come forward,” Pascal says. “That is a really personal decision left to the individual.”

Dale comments that once an FFP misstep reaches publication stage, whistleblowers who are co-authors must also step forward or be found complicit later.

“It's an honour system,” says Mortimer Litt, associate dean and senior scientific investigator at Harvard Medical School. “We do not have thought police monitoring every lab notebook every day. We depend on whistleblowers.”

Stepping into that role is not easy. DeNino found that out when he was counter-accused of planting the falsified data. “It's not something I would suggest doing if you don't have a strong character,” he says. “You can never predict what will happen.”

It's a roll of the dice — but so is remaining quiet, says Turens. “You may choose to ignore it, but look at it from a selfish standpoint. If the person is ever caught, your career will have a spot on it.” ■

Kendall Powell is a freelance science writer in Broomfield, Colorado.

*Some names have been changed to protect privacy.

WEB LINK

Office of Research Integrity
<http://ori.dhhs.gov>

CSI: HARVARD

When an allegation of scientific misconduct is made at Harvard Medical School, Mortimer Litt, senior scientific investigator, springs into action. His investigation may not be drama-filled like forensics television shows, but it does involve reconstructing what happened.

“In academia, we value nothing more than accurate reporting. We pursue it to the bitter end,” he says. He and colleagues go to the lab to secure original data and lab notebooks so they can't be altered. Litt does a preliminary investigation and reports his factual findings to a panel of three senior professors.

This panel determines if there is enough “presumptive evidence of misconduct” to send it to the larger standing committee for investigations. That committee delivers a decision and recommendations for sanctions, if any, to the accused's dean and to the funding agency, if the case involved grants.

Litt says the process is not a trial, but a search for the truth. The accused gets copies of all notes and testimony; nothing is sprung on them Perry Mason-style. The process can take from a couple of months to six years. Sometimes lab work continues, but sometimes the accused is banned from the lab until the matter is resolved.

Litt enjoys using his intellect to solve such mysteries — comparing duplicated images or forged signatures. “I always get a thrill when I make connections,” he says. “Then, my next thought is that this is a human tragedy.”

K.P.

MOVERS

William Harris, president and chief executive, Science Foundation Arizona



2001-06: Director-general, Science Foundation Ireland, Dublin

2000-02: Vice-president for research, University of South Carolina, Columbia

1996-2000: President and executive director, Columbia University Biosphere 2 Center, Tucson, Arizona; and associate director, Earth Institute, Columbia University, New York

It's not often that a scientific career reaches beyond one's own discipline. But William Harris has helped to transform scientific infrastructures around the world: a commitment to science and education is the cornerstone of his career.

A chemistry professor at William and Mary College in Williamsburg, Virginia, first brought the subject to life for Harris, who went on to do graduate work in spectroscopy at the University of South Carolina (USC) and then to teach at Furman University in Greenville, South Carolina. He learnt about biology on a sabbatical at the National Institutes of Health, where he used spectroscopy with biophysicist Ira Levin. A sabbatical at the National Science Foundation (NSF) led to a job as programme director in physical chemistry.

By the end of his 18 years at the NSF, Harris was director of mathematical and physical sciences, overseeing a \$750-million annual budget. In addition to establishing 24 science and technology centres to support interdisciplinary research, his greatest satisfaction came from developing the research programme for undergraduates in the chemistry division — which became a standard throughout the NSF divisions. "I became a person who could help gifted scientists do their science better," says Harris of his ability to remove barriers to scientific work.

Although he calls his position at the NSF "the best job I ever had", Columbia University offered an opportunity that he couldn't refuse — to head Biosphere 2, an enclosed environmental system that aroused controversy about how self-contained it really was. Harris's interest in climate problems was fired by Columbia's ambitious agenda for its Earth Institute experiment. "It was a challenge to do something that hadn't been done before — and make an important education statement in the process," he says.

After a brief stint back at USC, as vice-president for research, Harris was head-hunted to lead Science Foundation Ireland. In five short years, he's built Ireland's research — engaging its universities in previously uncultivated areas such as biotechnology, and attracting distinguished scientists from the Max Planck Institutes and Bell Labs to set up research centres. He has been asked to advise the European Commission, China and India on developing their scientific economics.

But it is to Arizona that Harris moves next. He will soon head Science Foundation Arizona, a new non-profit institute dedicated to doing for Arizona what he has done for Ireland — build the new research programmes from the ground. ■
Virginia Gewin

SCIENTISTS & SOCIETIES

Postdoc prep

Postdocs in the United States have never been held in the same level of esteem as other professionals. Members of under-represented groups, such as women and racial minorities, face additional challenges in the scientific workforce, including feelings of isolation, cultural misunderstandings and difficulties in communicating need. Mentoring and support from faculty members, staff and peers have been cited as key factors in ensuring postdoctoral success.

The National Postdoctoral Association (NPA) is committed to serving the diverse needs of the entire postdoctoral community. We lead the Diversity Committee of the NPA, which focuses on advocacy and providing resources for members of racial/ethnic minority groups, women, people with disabilities, and those of different sexual orientation. The committee has been very active since its inception three years ago, including organizing workshops at the annual meetings of the NPA, the Society for the Advancement of Chicanos and Native Americans in Science, and the Compact for Faculty Diversity's Institute on Teaching and Mentoring.

In partnership with the Alliance for Graduate Education and the Professoriate Program of Howard University (in Washington DC) and the University of

Texas at El Paso, we have created a unique programme called the Institute on Preparing for the Postdoctorate in Science, Technology, Engineering and Mathematics. This two-day workshop seeks to help students nearing the end of graduate school to understand the importance of postdoctoral training and the steps necessary to ensure a successful experience. About 40 students took part in each of the past two years.

Topics of discussion include different types of postdoctoral opportunities, how to select the best institution and lab, how to secure funding and grants, developing a career plan, and the unique experiences of under-represented groups. Speakers include former and current postdoctoral fellows, faculty members interested in hiring postdocs and representatives from funding agencies. The key features of the workshop are that it is informal and interactive. We have received positive feedback from participants and speakers. We will be expanding this year by providing similar programmes at other organizations with postdocs and at scientific society meetings. ■

Jabbar Bennett is chair and Cherie Butts is vice-chair of the NPA's Diversity Committee.

► www.nationalpostdoc.org/committees/diversity_committee

GRADUATE JOURNAL

On Mozart's wavelength

One of the most rewarding aspects of being a graduate student is being sent halfway around the world to present your research and interact with other scientists at scientific meetings. The distance travelled is not important, but it is always nice to combine the professional experience with explorations of a new city, country, continent and culture. As I write this entry, I have just attended the European Geosciences Union meeting in Vienna, Austria. Although Vienna is located far away from any ocean, what can go wrong when you have unlimited access to Wiener schnitzel, apple strudel and Mozart?

At the first conference I attended, I was scheduled to present my research as one of the last talks of the entire conference. An impressive crowd of about a dozen showed up, including the organizers and four fellow students from Hawaii! On my second appearance, I was scheduled to speak just before the legendary oceanographer Wally Broecker, and a steady flow of people filled the room until it was announced that Wally's talk was cancelled owing to illness, and the majority left. Luckily, I have not spoken to empty seats since then. Each conference has pushed me closer to graduation, articles have been solicited and published, collaborations initiated, postdoc positions offered and friends made all over the world. In addition, I have enjoyed apple strudel to the tunes of Mozart in the place where they were composed more than two centuries ago. ■

Andreas Andersson is a final-year PhD student in oceanography at the University of Hawaii.

Taking good care of myself

Someone to watch over me.

Ian R. MacLeod

The social worker came a day or so before I arrived. He was as briskly pleasant as the occasion, which I'd long been dreading, allowed. He was dressed bizarrely, but people from the future always are.

"We'll need to send a few helpful machines back with you," he murmured as he inspected our spare bedroom and the bathroom and then the kitchen, which no doubt looked ridiculously primitive to him. "But nothing that'll get in your way."

Helen was equally reassuring when she came home that evening. "It's a tremendous challenge," she told me. "You always say you like challenges."

"I mean stuff like climbing, hang-gliding, pushing things to the edge, not looking after some senile version of myself"

"Josh." She gave me one of her looks. "You have no choice."

She was right — and there was plenty of space in our nice house. It was as if we'd always planned on doing precisely this, although I hated the very thought.

I arrived a couple of mornings after, flanked by swish-looking machines, although I was just as pale and dithery as I'd feared. The creature I'd eventually become couldn't walk, could barely see, and certainly didn't comprehend what was happening to him. Exactly how long, I wondered (and secretly hoped), can I possibly last like this?

Scampering around our house like chromium shadows, the machines performed many of the more obvious and unpleasant duties, but much was still expected of me. I had to sit and talk, although my elderly self rarely said anything in response, and none of it was coherent. I also had to help myself eat, and wipe away the spilled drool afterwards. I had to hold my own withered hand.

"Do you remember this house — I mean, you must have lived here?"

But I was much too far gone to understand. Not, perhaps, in a vegetative state yet, but stale meat at very best.

Sometimes, I took me out, pretending to push the clever chair which was in fact

more than capable of doing everything — except getting rid of this cadaverous ghost — by itself. My work suffered. So did my relationship with Helen. I joined a self-help group. I sat in meeting halls filled with other unfortunates who'd had the care of their future selves foisted on them. We debated in slow circles why our future children,



or the intelligences that perhaps governed them, had seen fit to make us do this. Were they punishing us for the mess we'd made of their world? Or were these addled creatures, with their lost minds, their failed memories, their thin grip on this or any other kind of reality, somehow the means of achieving time travel itself? Predictably, various means of killing were discussed, from quiet euthanasia to violent stabbings and cliff-top falls. But that was the thing: complain as we might, not one of us ever seemed capable of harming ourselves. Not, anyway, the selves we would eventually become.

I declined. The machines, with a will of their own, grew yet more sophisticated and crouched permanently beside me as I lay immovably in my spare bedroom cradled in steel pipes and crystal insertions.

They fed me fresh blood, fresh air. I doubted if this husk I'd become was conscious of any presence other than its own dim existence now, but still I found myself sitting beside me, and talking endlessly about things I couldn't remember afterwards. It was as if I was trapped in a trance, or that part of me was dying as well. I lay entirely naked now under clever sheets that cleaned themselves. Occasionally, inevitably, I would lift them up, and breathe the stale air of my own mortality, and study the thin limbs and puckered flesh of what I would eventually become. The death itself was surprisingly easy. The machines saw to it that there was no pain, and I was there; I made sure I didn't die alone. A faint rattle, a tiny spasm. You're left wondering what all the fuss is about.

After the funeral, which of course I also had to arrange myself, and was far more poorly attended than I might have hoped, and then the scattering of my ashes at the windy lip of one of my favourite climbs, I looked around at my life like a sleeper awakening. Helen had left me, although quietly, without fuss. My house felt empty, but I knew that it was more to do with that old man than with her. I'm back to climbing regularly now. I'm back to freefall and hang-gliding. I find that I enjoy these sports, and many other kinds of dangerous and challenging physical activities, even more. After all, I know they can't kill me, and that the last phase of my life really isn't so very bad. But things have changed, for all that, and I still sometimes find myself sitting alone in my spare bedroom gazing at the taut sheets of that empty bed, although I and all those future machines have long gone. The sad fact is, I miss myself dreadfully, now that I'm no longer here. ■

Ian R. MacLeod's latest novel, *The House of Storms*, is just out in paperback, and a short-story collection, *Past Magic*, will also be published shortly. He's just completed a new novel, *Song of Time*, which looks back at this current century through the eyes of elderly woman contemplating a leap into virtual intelligence. He lives in the river town of Bewdley.

JACEY